Layer 3 Industrial Ethernet Switch

CLI User Manual

Document Version: 01

Issue Date: 12/07/2021

Preface

Switch CLI user manual has introduced:

- CLI configuration interface login
- CLI configuration rule and method
- Network management functions related CLI introduction

Audience

This manual applies to the following engineers:

- Network administrators
- Technical support engineers
- Network engineer

Text Format Convention

Format	Description
" "	Words with “” represent the interface words. Fox example “Port number”.
>	Multi-level path is separated by ">". Such as opening the local connection path description: Open "Control Panel> Network Connection> Local Area Connection".
Light Blue Font	It represents the words clicked to achieve hyperlink. The font color is as follows: ‘Light Blue’.
About this chapter	The section ‘about this chapter’ provide links to various sections of this chapter, as well as links to the Principles Operations Section of this chapter.

Symbols

Format	Description
Notice	Remind the announcements in the operation, improper operation may result in data loss or equipment damage.
Warning	Pay attention to the notes on the mark, improper operation may cause personal injury.

Format	Description
Notes	Make a necessary supplementary instruction for operation description.
Key	Configuration, operation, or tips for device usage.
Tips	Pay attention to the operation or information to ensure success device configuration or normal working.

Port Convention

The port number in this manual is only an example, and does not represent the actual port with this number on the device. In actual use, the port number existing on the device shall prevail.

Revision Record

Version No.	Date	Revision note
01	2021-12-07	Product release

CONTENTS

PREFACE 1
CONTENTS 1
1 LOGIN THE SWITCH CONFIGURATION 18
1.1 LOGIN THE SWITCH FUNCTION OVERVIEW 18
1.2 LOGIN TO THE SWITCH 18
1.2.1 Login to the Switch via Serial Port 18
1.2.2 Login the Switch via Telnet 20
1.2.3 Login to the Switch via SSH 22
1.2.4 Login the Switch via WEB: 24
1.2.5 Manage the switch via Network Management Software 25
1.3 COMMAND LINE 25
1.3.1 Command Analysis 25
1.3.2 Command Line Mode 26
1.3.3 Shortcut Key 26
1.4 COMMON COMMAND 28
1.4.1 Password Verification 28
1.4.2 Customization Display 28
1.4.3 Configuration Management 29
1.4.4 System Upgrading 30
1.4.5 Debug Mode 30
2 USER CONFIGURATION 31
2.1 ADD USER 31
2.2 DELETE USER 33
2.3 VIEW CURRENT ONLINE USERS 33
2.4 CONSOLE LOGIN MANAGEMENT 34
2.5 VIRTUAL TERMINAL LOGIN MANAGEMENT 35
2.6 TIMEOUT LOGOUT 35
3 PORT CONFIGURATION 37
3.1 ENTER PORT CONFIGURATION MODE 37
3.2 PORT RATE LIMIT 38
3.3 PORT SETTINGS 39
3.3.1 Combo PortTransmission Media 39
3.3.2 Duplex Mode 40
3.3.3 Flow Control 41

3.3.4 Max-Frame 41
3.3.5 Interface Switch 42
3.3.6 Rate 43
3.4 PORT ISOLATION 44
3.5 STORM SUPPRESSION 45
3.6 MAC ADDRESS 46
3.6.1 Clear Dynamic MAC address 46
3.6.2 MAC Address Learning 47
3.6.3 MAC Address Aging-Time 47
3.6.4 Static MAC Address Filtering 48
3.6.5 Multicast MAC Address Filtering 49
3.6.6 Display MAC Address Table 50
3.7 MIRROR COMMAND 51
3.7.1 Port Mirror Configuration 51
3.7.2 Delete Port Mirror 52
3.8 LINK AGGREGATION CONFIGURATION 52
3.8.1 Dynamic Aggregation System Priority 52
3.8.2 Dynamic Aggregation Port Priority 53
3.8.3 Dynamic Aggregation Port Timeout 54
3.8.4 Add Dynamic Aggregation Group 54
3.8.5 Add static LACP 55
3.8.6 Link Aggregation Load Balance Mode 56
3.8.7 Displays Dynamic Aggregation Group 57
3.8.8 Displays Static Aggregation Group 57
3.9 AGGREGATION PROTECTION 58
3.9.1 Enable Aggregation Protection 58
3.9.2 Displays Aggregation Protection Status Information 59
3.10 PORT STATISTICS 60
3.10.1 Display Port 60
3.11 LINK FLAPPING PROTECTION CONFIGURATION 62
3.11.1 Enable Link Flapping Protection 62
3.11.2 Enable Link Flapping Auto-Recovery 62
3.11.3 Configure Recovery Interval of Link Flapping 63
3.11.4 Configure Detection Interval of Link Flapping 64
3.11.5 Configure Time Threshold Value of Link Flapping 64
3.11.6 Check Link Flapping Protection Configuration 65
4 VLAN CONFIGURATION 66
4.1 VLAN OVERVIEW 66
4.2 PRINCIPLE DESCRIPTION 67
4.2.1 VLAN Tags 67
4.2.2 Link and Interface Types 70
4.2.3 Default VLAN 71
4.2.4 Adding and Removing VLAN Tags 72

4.2.5 VLAN Division 74
4.2.6 Intra-VLAN Communication 77
4.2.7 Inter-VLAN Communication 78
4.2.8 Intra-VLAN Layer 2 Isolation 79
4.2.9 mVLAN 79
4.3 CONFIGURE VLAN 80
4.3.1 Enter VLAN Configuration Mode 80
4.3.2 Add VLAN ID 80
4.3.3 Port Type 81
4.3.4 Port Default VLAN 82
4.3.5 Classify VLAN Based on Port 83
4.3.6 Port Receive Frame Type 84
4.3.7 Port Entry Filtering 85
4.3.8 VLAN Division Based on Subnet/MAC/ Protocol 85
4.3.9 Configure VLAN Classification Group 88
4.3.10 Configure the Interface VLAN Classification Group 89
4.3.11 Display VLAN Information 89
5 RING CONFIGURATION 91
5.1 RING OVERVIEW 91
5.2 PRINCIPLE DESCRIPTION 92
5.2.1 Network ID 92
5.2.2 Ring port 92
5.2.3 Ring Type 92
5.2.4 Master/Slave Mode 94
5.3 RING CONFIGURATION 95
5.3.1 Global Ring Enablement 95
5.3.2 Create Ring NetworkGroup 96
5.3.3 Display Ring Network Information 97
6 STP/RSTP/MSTP CONFIGURATION 98
6.1 STP/RSTP/MSTP OVERVIEW 98
6.1.1 STP/RSTP Overview 98
6.1.2 DHCP Overview 99
6.2 PRINCIPLE DESCRIPTION 99
6.2.1 STP Principle Description 99
6.2.2 RSTP principle description 112
6.2.3 MSTP Principle Description 117
6.3 CONFIGURE STP/RSTP/MSTP 126
6.3.1 Global Spanning-tree Enablement 126
6.3.2 MSTP Instance Configuration 126
6.3.3 Bridge Configuration 129
6.3.4 Port Configuration 134
6.3.5 Instance Port Configuration 145
6.3.6 Display Spanning Tree Information 147

7 ERPS CONFIGURATION 154
7.1 ERPS OVERVIEW 154
7.2 PRINCIPLE DESCRIPTION 154
7.2.1 Basic ERPS Concepts 154
7.2.2 RAPS PDUs 161
7.2.3 ERPS Operation Mechanism 163
7.3 CONFIGURE ERPS 167
7.3.1 Timer Configuration 167
7.3.2 ERPS Ring Configuration 171
7.3.3 ERPS Instance Configuration 174
8 LOOP DETECTION CONFIGURATION 185
8.1 OVERVIEW 185
8.2 PRINCIPLES 185
8.3 GLOBAL ENABLEMENT CONFIGURATION 186
8.4 FORCES THE PORT CLOSED BY THE PROTOCOL TO OPEN 187
8.5 CONFIGURE PROTECT VLAN 187
8.6 CONFIGURE THE PORT RECOVERY TIME 188
8.7 CONFIGURE THE PROBE PACKET INTERVAL 188
8.8 DISPLAYS LOOP DETECTION INFORMATION 189
9 IGMP CONFIGURATION 191
9.1 OVERVIEW 191
9.2 PRINCIPLES 191
9.2.1 IGMPv1 Fundamentals 192
9.2.2 IGMPv2 Fundamentals 194
9.2.3 IGMPv3 Fundamentals 196
9.2.4 IGMP SSM Mapping 200
9.2.5 IGMP Proxy 201
9.3 IGMP CONFIGURATION 204
9.3.1 Configure IGMP Basic Functions 204
9.3.2 Adjust IGMP Performance 210
9.3.3 Configure IGMP Limit 219
9.3.4 Configure IGMP SSM Mapping 220
10 IGMP SNOOPING CONFIGURATION 223
10.1 OVERVIEW 223
10.2 PRINCIPLE DESCRIPTION 223
10.2.1 IGMP Snooping Relative Ports 224
10.2.2 Implementation 225
10.3 CONFIGURE IGMP SNOOPING 227
10.3.1 IGMP Snooping Enablement 227
10.3.2 IGMP Snooping Querier Enablement 228
10.3.3 IGMP Snooping Port Fast-leave Enablement 229
10.3.4 IGMP SnoopingPort Suppression Enablement 230
10.3.5 Configure the Routing Interface for IGMP Snooping Multicast Group 230

10.3.6 Configure IGMP Snooping Multicast Permanent Group 231
10.3.7 Configure IGMP Snooping to Send the Source IP Address 232
10.3.8 Display the IGMP Snooping Multicast Group Routing Interface 232
10.3.9 Display IGMP SnoopingMulticast Statistics 233
10.3.10Display IGMP Snooping Multicast Group Information 233
11 GMRP AND MMRP CONFIGURATION 235
11.1 OVERVIEW 235
11.2 PRINCIPLE DESCRIPTION 236
11.2.1 GARP 236
11.2.2 MRP 241
11.3 CONFIGURE GMRP OR MMRP 247
11.3.1 GlobalGMRP or MMRP Enablement 247
11.3.2 Port GMRP or MMRP Enablement 247
11.3.3 GMRP or MMRP Registration Mode 248
11.3.4 GMRP or MMRP Timer 249
11.3.5 Display GMRP or MMRP Configuration Information 249
11.3.6 Display GMRP or MMRP State Machine Information 250
11.3.7 Display GMRP or MMRP Message Statistics 250
11.3.8 Display GMRP or MMRP Timer Information 251
12 GVRP AND MVRP CONFIGURATION 252
12.1 OVERVIEW 252
12.2 PRINCIPLE DESCRIPTION 253
12.2.1 GARP 253
12.2.2 MRP 258
12.3 CONFIGURE GVRP AND MVRP 263
12.3.1 Global GVRP or MVRP Enablement 263
12.3.2 GVRP or MVRP Dynamic VLAN Enablement 264
12.3.3 Port GVRP or MVRP Enablement 265
12.3.4 GVRP or MVRP Registration Mode 265
12.3.5 GVRP or MVRP Timer 266
12.3.6 Display Dynamic VLAN Information 267
12.3.7 Display GVRP or MVRP Configuration Information 267
12.3.8 Display GVRP or MVRP State Machine Information 268
12.3.9 Display GVRP or MVRP Message Statistics 268
12.3.10Display GVRP or MVRP Timer Information 269
13 PIM-DM CONFIGURATION 270
13.1 OVERVIEW 270
13.2 PRINCIPLE DESCRIPTION 271
13.2.1 Concepts 271
13.2.2 PIM-DM 273
13.3 CONFIGURE PIM-DM 277
13.3.1 Enable PIM-DM 277
13.3.2 PIM-DM DR Priority 278

13.3.3 The PIM - DM GenID Information 278
13.3.4 PIM-DM Neighbor Reachable State Time 279
13.3.5 Time Interval of PIM-DM Hello Message 280
13.3.6 PIM-DM Neighbor Filter 280
13.3.7 IPM-DM Status Refresh Message Enablement 281
13.3.8 The Time Interval During Which PIM-DM Sends a Status Refresh Message 282
13.3.9 The Time Interval Between PIM-DM Receiving a Status Refresh Message 283
13.3.10 Display PIM-DM Detail Information 283
13.3.11 Display PIM-DM Interface Information 284
13.3.12 Displays the Local Multicast Group Members of the PIM-DM 285
13.3.13 Displays PIM-DM Multicast Routing Table Entries 286
13.3.14 Display PIM-DM Neighbor Information 286
13.3.15 Displays PIM-DM Next Hop Information 287
14 PIM-SM CONFIGURATION 288
14.1 OVERVIEW 288
14.2 PRINCIPLE DESCRIPTION 289
14.2.1 Concepts 289
14.2.2 PIM-SM (ASM Model) 291
14.2.3 PIM-SM (SSM Model) 298
14.3 CONFIGURE PIM-SM 301
14.3.1 Enable PIM-SM 301
14.3.2 PIM-SM DR Priority 302
14.3.3 PIM-SM GenID Information 303
14.3.4 PIM-SM Neighbor Reachable State Time 303
14.3.5 Time Interval of PIM-SM Hello Message 304
14.3.6 PIM-SM Neighbor Filter 305
14.3.7 Restrictions on Illegal Registration of Messages in PIM-SM 306
14.3.8 PIM C-BSR 307
14.3.9 PIM Registration Message Validation and Compatibility With Cisco Standards 307
14.3.10 PIM C-RP is compatible with Cisco standards 308
14.3.11 PIM BSR Ignores C-RP Priority 309
14.3.12 The Time Interval Between PIM Sending and/or Pruning Messages 310
14.3.13 The Rate at Which PIM Receives and Processes Multicast Service Messages 310
14.3.14 PIM Checks the Accessibility of the RP 311
14.3.15 The VLAN interface or source IP address where the PIM sends the registered message 312
14.3.16 PIM Registration Suppression Time 312
14.3.17 PIM Static PR 313
14.3.18 PIM C-RP 314
14.3.19 PIM KAT Timer Aging Time 315
14.3.20 PIM SPT Switch 316
14.3.21 Displays BSR Running Information for PIM-SM 317

14.3.22Display PIM-SM Interface Information 317
14.3.23Displays the Local Multicast Group Members of the PIM-SM 319
14.3.24Displays the PIM-SM Multicast Routing Table Entry 319
14.3.25Display PIM-SM Neighbor Information 321
14.3.26Display PIM-SM Next Hop Information 321
14.3.27Display PIM-SM RP Information 322
14.3.28Displays the RP Address of the PIM-SM Multicast Group 323
15 VRRP CONFIGURATION **324**
15.1 OVERVIEW 324
15.2 PRINCIPLE DESCRIPTION 325
15.2.1 Concepts 325
15.2.2 VRRPv2 and VRRPv3 Advertisement Packets 326
15.2.3 VRRP Implementation 330
15.2.4 VRRP Active/Standby 334
15.2.5 VRRP Load Balancing 335
15.2.6 VRRP Monitoring the Uplink Status 336
15.3 CONFIGURE VRRP 336
15.3.1 Create VRRP Group and Enter VRRP Configuration View 336
15.3.2 VRRP Interface Enablement 337
15.3.3 VRRP Virtual IP Address 337
15.3.4 Preemption Mode Enablement 338
15.3.5 Preemption Delay Time 338
15.3.6 Priority of VRRP Device 339
15.3.7 VRRP Notification Message Interval 340
15.3.8 VRRP Message Authentication Mode 341
15.3.9 Authentication Word of VRRP Message 342
15.3.10Configure the Monitoring Specified Interface 343
15.3.11Track IPDT Session 344
15.3.12Enable VRRP Backup Group 344
15.3.13Disable VRRP Backup Group 345
15.3.14Display VRRP Backup Group Information 346
16 RIP CONFIGURATION **347**
16.1 OVERVIEW 347
16.2 PRINCIPLES 347
16.2.1 RIP Principles 347
16.2.2 RIP-2 Enhanced Features 350
16.2.3 Split Horizon and Poison Reverse 352
16.3 CONFIGURE RIP 353
16.3.1 Start RIP Process 353
16.3.2 Enable RIP in the Specified Network Segment 354
16.3.3 Configure IP Address of RIP Neighbor in NBMA Network 355
16.3.4 Add Static RIP Route 356
16.3.5 Add Default Routing to RIP Routing Database 356

16.3.6 Default Route Metric 357
16.3.7 RIP Route Management Distance 358
16.3.8 Access List Route Filtering 358
16.3.9 Other Routing Protocols Route Import 359
16.3.10 Block RIP Broadcast 360
16.3.11 Time of RIP Timer 360
16.3.12 RIP Version 361
16.3.13 Maximum Number of RIP Route 362
16.3.14 RIP Routing Measures Offset 363
16.3.15 RIP Route UDP to Receive Cache Size 363
16.3.16 RIP Message Authentication Mode 364
16.3.17 RIP Message Authentication Key Chain 365
16.3.18 RIP Message Authentication Password 366
16.3.19 Receive RIP Message Enablement 366
16.3.20 Accept Message of Specified RIP Version 367
16.3.21 Send RIP Message Enablement 368
16.3.22 Send the Message of the Specified RIP Version 368
16.3.23 RIP Horizontal Split Enablement 369
16.3.24 Display Routing Information learned by RIP 370
16.3.25 Display the Routing Information in the RIP Routing Information Base 371
16.3.26 Display RIP Interface Information 373
17 OSPF CONFIGURATION 375
17.1 OVERVIEW 375
17.2 PRINCIPLE DESCRIPTION 376
17.2.1 OSPF Characteristics 376
17.2.2 OSPF Mechanism 377
17.2.3 OSPF Packet Types 380
17.2.4 OSPF Network Types 381
17.2.5 DR/BDR Election 382
17.2.6 OSPF State Machine 386
17.2.7 OSPF Adjacency Establishment 391
17.2.8 OSPF Area 392
17.2.9 OSPF Fast Convergence 399
17.2.10 OSPF Virtual Link 400
17.2.11 OSPF Route Summarization and Filtering 401
17.2.12 OSPF Multi-Process 403
17.2.13 OSPF RFC 1583 Compatibility 403
17.3 OSPF CONFIGURATION 404
17.3.1 Enter OSPF Router Configuration View 404
17.3.2 OSPF Route ID 404
17.3.3 OSPF Area and Enablement 405
17.3.4 Configure and Publish a Host Route 406
17.3.5 Configure Neighbor Interface Address in the NBMA Network 406

17.3.6 Create Virtual Connection 407
17.3.7 Routing Within the Aggregation Area 408
17.3.8 Type-3 LSA Filter 409
17.3.9 Stub Area 410
17.3.10 NSSA Area 411
17.3.11 Area Shortcut Mode 412
17.3.12 Default Costs for Introduced Default Routes 413
17.3.13 Configure Aggregation of External Routes and Notify or Suppress 414
17.3.14 Link Cost Bandwidth Reference Value 414
17.3.15 Opaque LSA Publishing and Receiving Capabilities 415
17.3.16 Compatibility with RFC1583 416
17.3.17 Introduce Default Route 417
17.3.18 OSPF Routing Default Metric 418
17.3.19 OSPF Route Management Distance 418
17.3.20 Ingress Route Information Filter 419
17.3.21 Egress Route Information Filter 420
17.3.22 Introduce Additional Routing Information 421
17.3.23 Interface Hello Message Control 422
17.3.24 SPF Timer 422
17.3.25 The Maximum Number of LSA 423
17.3.26 Maximum Concurrent Number of Current DD Packet 424
17.3.27 Regional OSPF Message Validation 424
17.3.28 OSPF ABR Type 425
17.3.29 Interface OSPF Message Validation 426
17.3.30 Interface OSPF Message Authentication Key 427
17.3.31 Interface OSPF Cost 428
17.3.32 Interface LSA Database Filter 429
17.3.33 Interface OSPF Neighbor Failure Time 429
17.3.34 Time Interval of Sending Hello Message 430
17.3.35 Interface Retransmission LSA Interval 431
17.3.36 Interface to LSA Transmission Delay Time 431
17.3.37 Disable the OSPF Function of the Interface 432
17.3.38 Interface OSPF MTU Value 432
17.3.39 Interface DD Exchanges Ignore MTU 433
17.3.40 Interface OSPF Network Type 434
17.3.41 The DR Priority of the Interface 434
17.3.42 Display OSPF Route Information 435
17.3.43 Display OSPF Database Information 436
17.3.44 Display OSPF Neighbor Information 438
18 BGP CONFIGURATION 440
18.1 OVERVIEW 440
18.2 PRINCIPLES 441
18.2.1 BGP Concepts 441

18.2.2 BGP Working Principles 443
18.2.3 Interaction Between BGP and IGPs 446
18.2.4 BGP Route Selection Rules and Load Balancing 447
18.2.5 Route Reflection 452
18.2.6 BGP Confederation 454
18.2.7 Route Summarization 455
18.2.8 Route Dampening 456
18.3 BGP CONFIGURATION 457
18.3.1 Enter BGP Route Configuration View 457
18.3.2 Enter the IPv4 or IPv6 Address Family 458
18.3.3 BGP Route Aggregation Address 458
18.3.4 BGP Always Comparison MED Value 459
18.3.5 The BGP optimal Path Ignore the AS-Path Length 459
18.3.6 Comparison of AS-Path Length of BGP Optimum Path 460
18.3.7 BGP Optimum Path Routing ID Comparison 461
18.3.8 MED Property of BGP Optimum Path 461
18.3.9 BGP Route Reflection 462
18.3.10 Configure the Group Identifier for the Routing Reflector 463
18.3.11 Configure the Identifier for the AS Alliance 463
18.3.12 Configure the Member AS for the AS Alliance 464
18.3.13 EBGP Route Suppression 465
18.3.14 Set the BGP Address Family to IPv4 Unicast by Default 467
18.3.15 The Default Local Priority Value for BGP 468
18.3.16 Compare MED Values of the Same AS Counterparts 468
18.3.17 Force the AS-Path of the First Location to be the EBGP Route 469
18.3.18 Reset the Direct Connection EBGP Fault Interface 470
18.3.19 Record BGP Status Change Information 470
18.3.20 BGP Route ID 471
18.3.21 BGP Next Hop Detection Interval 471
18.3.22 Specifies the Management Distance for the Routing Prefix 472
18.3.23 Managed Distances for Internal/External/Local BGP Routes 473
18.3.24 Exit the BGP IPv4 or IPv6 Address Family 473
18.3.25 Network Segment Routing Notification BGP Routing Table 474
18.3.26 Route Synchronization Notify Network Information 475
18.3.27 Configure BGP Peers/Peer Groups 476
18.3.28 Activate the Specified Peer/Peer Group 476
18.3.29 The Time Interval Between Sending the BGP Route Update Message 477
18.3.30 The Number of Occurrences of the Same AS in the Neighbor Route AS List 478
18.3.31 The Time Interval Between Sending the Local Originating BGP Route 479
18.3.32 The Routing Properties Associated with the Transport Neighbors Remain Unchanged 480
18.3.33 Dynamic Updates and Routing Refreshes with Neighbors 481
18.3.34 Egress Route Filter With Neighbor 482

18.3.35TCP Connection Collision Detection with Neighbor 483
18.3.36Send a Default Route to a Peer/Peer Group 483
18.3.37Peer/Peer Group Description Information 484
18.3.38Filtering Strategy Based on ACL 485
18.3.39No Capacity Negotiation when Establishing a Connection 486
18.3.40Allows Non-directly Connected Neighbors to Establish EBGP Sessions 487
18.3.41Force Direct Contact with Neighbors for Multiple Hop 488
18.3.42AS-Path Access List Filtering 488
18.3.43Configure the Interface to the BGP Peer/Peer Group 489
18.3.44Number of Received Peer Prefixes 490
18.3.45Force Next-hop Addresses as Neighbors 491
18.3.46Ignore the Results of Peer Capacity Negotiation 492
18.3.47The Peer Connection is Passive 493
18.3.48Create a BGP Peer Group 493
18.3.49Add a Peer Group Member 494
18.3.50TCP Port Number for Communicating with Neighbors 495
18.3.51Routing Filtering Strategy based on IP Prefix List 495
18.3.52BGP Update Message Removes Private AS 496
18.3.53Route Mapping Strategy 497
18.3.54Route Reflection 498
18.3.55Router Server 498
18.3.56Community Property Transfer 499
18.3.57Disable BGP Connection 500
18.3.58Store BGP Original Routing Information 501
18.3.59Strict Ability Matching Connection 502
18.3.60BGP Session Timer with the Specified Peer 503
18.3.61Unsuppress the Routing Map 504
18.3.62Specify the BGP Connection Interface 504
18.3.63Configure the BGP Version of the Peer 505
18.3.64Configure the Weight Value of the Peer 506
18.3.65Introduce Additional Routing Information to the BGP 507
18.3.66BGP Connection Survival Interval and Holdtime 507
18.3.67Display BGP Route Information 508
18.3.68Display BGP IPv4 Unicast Route Information 511
18.3.69Display BGP Summary Information 514
19 IPV6 CONFIGURATION 517
19.1 OVERVIEW 517
19.2 PRINCIPLE DESCRIPTION 520
19.2.1 IPV6 Address 520
19.2.2 IPv6 Packet Format 526
19.2.3 ICMPv6 531
19.2.4 Neighbor Discovery 534
19.2.5 Static Routing 538

19.3 IPV6 CONFIGURATION 541
19.3.1 Create Layer 3 Interface 541
19.3.2 IPV6 Address 541
19.3.3 Static IPv6Route 542
19.3.4 Configure RA Message Related Parameters 542
19.3.5 The Maximum Transmission Unit 550
20 NAT CONFIGURATION 551
20.1 OVERVIEW 551
20.2 NAT CONFIGURATION 552
20.2.1 Configure NAT rule 552
20.2.2 Configure NAT Destination Network 553
20.2.3 Configure NAT Binding Port 554
20.2.4 Delete NAT Rules and Bound Ports 554
20.2.5 Display NAT Configuration 555
20.2.6 Display Activated NAT ACL Configuration 555
21 DHCP CONFIGURATION 557
21.1 OVERVIEW 557
21.2 PRINCIPLE DESCRIPTION 558
21.2.1 Network Elements in DHCP 558
21.2.2 DHCP Leases and Address Pools 559
21.2.3 DHCP Messages 561
21.2.4 The DHCP Server Assigns a Network Address to the Client That Accesses for the First Time 570
21.2.5 DHCP Client Reuses Network Address 574
21.2.6 DHCP Client Renews Its IP Address Lease 575
21.3 DHCP CONFIGURATION 577
21.3.1 Global DHCP Service Enablement 577
21.3.2 Enable Interface DHCP Relay 577
21.3.3 InterfaceDHCP Relay Server Address 578
21.3.4 DHCP Option82 Enablement 579
21.3.5 Treatment Strategy of DHCP Option82 580
21.3.6 Relay Identity of DHCP Option82 581
21.3.7 Remote Identity of DHCP Option82 582
21.3.8 Create DHCP Address Pool 583
21.3.9 DHCP Address Pool Subnet Segment 583
21.3.10Default Route of DHCP Address Pool 584
21.3.11DHCP Address Pool 585
21.3.12The Lease Time of DHCP Address Pool 586
21.3.13DNS Server Address 587
21.3.14Log Server Address 588
21.3.15WINS Server Address 589
21.3.16Display DHCP Information 589
22 DHCP-SNOOPING CONFIGURATION 594

22.1 OVERVIEW 594
22.2 PRINCIPLES 594
22.2.1 DHCP Snooping Principles 594
22.2.2 Option82 Supported by DHCP Snooping 596
22.3 CONFIGURE DHCP SNOOPING 598
22.3.1 Configure the Interface to Trust state 598
22.3.2 Configure the Rate of a DHCP Packet for the Layer 2 Port Interface 598
22.3.3 Enable Port Disable Duration 599
22.3.4 Configure Port Disable Duration 600
22.3.5 Delete DHCP Snooping security database entry 601
22.3.6 DHCP Snooping Enablement 601
22.3.7 VLAN DHCP Snooping Enable 602
22.3.8 Configure DHCP Snooping Static Binding Entries 603
22.3.9 Enable DHCP Snooping Check Function for MAC 603
22.3.10 Enable Option 82 Check for DHCP Snooping 604
22.3.11 Configure the Option 82 Handling Policy for DHCP Snooping 605
22.3.12 Configure the Option 82 Circuit-id of DHCP Snooping 606
22.3.13 Configure the Option 82 Remote-id of DHCP Snooping 606
22.3.14 Configure the Value of Option 82 Subscriber-id for DHCP Snooping 607
22.3.15 Display DHCP Snooping Information 608
23 SNMP CONFIGURATION 609
23.1 OVERVIEW 609
23.2 PRINCIPLES 610
23.2.1 SNMP Management Model 610
23.2.2 SNMPv1/SNMPv2c 613
23.2.3 SNMPv3 617
23.3 CONFIGURE SNMP 620
23.3.1 SNMP Enablement 620
23.3.2 SNMP View 621
23.3.3 SNMP Community Name 622
23.3.4 SNMP Group 623
23.3.5 SNMP User 624
23.3.6 SNMP Trap Destination 625
23.3.7 View information about SNMP 626
24 LLDP CONFIGURATION 628
24.1 OVERVIEW 628
24.2 PRINCIPLES 629
24.2.1 Working Principle 629
24.2.2 Message Structure 630
24.2.3 Message Transmission Mechanism 634
24.2.4 Networking Mode 635
24.3 CONFIGURE LLDP 637
24.3.1 LLDP Enablement 637

24.3.2 LLDP Port Operating Mode 638
24.3.3 Time Interval of Sending LLDP Message 638
24.3.4 LLDP Interface Management Address 639
24.3.5 Encapsulation Format of LLDP Message 640
24.3.6 Display LLDP Neighbor Information 640
24.3.7 Display LLDP Statistics Information 642
24.3.8 Display LLDP Local Information 643
24.3.9 Display LLDP Status Information 644
25 QOS CONFIGURATION 646
25.1 OVERVIEW 646
25.1.1 QoS Introduction 646
25.1.2 Priority Mapping 647
25.1.3 Flow Monitoring, Traffic Shaping and Interface Speed Limit 650
25.1.4 Congestion Avoidance and Congestion Management 650
25.2 QOS CONFIGURATION 654
25.2.1 Configure Global QOS Enable/Disable 654
25.2.2 Configure the Queue Bitmap 655
25.2.3 Configure Queue Scheduling Mode 656
25.2.4 Configure the DSCP-COS Bitmap 657
25.2.5 Configure DSCP -DSCP Bitmap 658
25.2.6 Create a CLASS-MAP 658
25.2.7 Create a POLICY-MAP 659
25.2.8 Configure the CLASS-MAP Property 660
25.2.9 Configure the POLICY-MAP Property 661
25.2.10 Configure the POLICY-MAP-C Property 661
25.2.11 Configure QOS Interface Mode 662
25.2.12 Display the Queue Bitmap 664
25.2.13 Display Queue Scheduling Mode 665
25.2.14 Display DSCP-COS Bitmap 665
25.2.15 Display DSCP- DSCP Bitmap 666
25.2.16 Display a CLASS-MAP 667
25.2.17 Display a POLICY-MAP 667
26 ACL CONFIGURATION 669
26.1 OVERVIEW 669
26.2 PRINCIPLES 670
26.2.1 ACL Principles 670
26.2.2 Classification of ACL 671
26.2.3 Common Matches of ACL 672
26.2.4 Effective Time Period of ACL 675
26.3 ACL CONFIGURATION 677
26.3.1 Configure IPv4 Extended ACL Based on IP Addresses 677
26.3.2 Configure IPv4 Extended ACL Based on IP Addresses 678
26.3.3 Configure Other IPv4 Protocol Extended ACL based on IP Addresses 679

26.3.4 Configure IPv4 TCP Extended ACL Based on IP Addresses 680
26.3.5 Configure IPv4 UDP Extended ACL Based on IP Addresses 682
26.3.6 Configure Character Type ACL Based on IPv4 Addresses 684
26.3.7 Configure Character Type Standard ACL Based on Ipv6 Addresses 687
26.3.8 Configure Character Type Extended ACL Based on Ipv6 Addresses 688
26.3.9 View All Configured ACL 691
26.3.10 Configure time-range 692
26.3.11 time-range Binds to the ACL 693
26.3.12 Activate IP ACL 694
26.3.13 Configure ACL based on MAC address 695
26.3.14 View all configured MAC ACL 696
26.3.15 Time-range and MAC ACL binding 696
26.3.16 Activate MAC ACL 697
26.3.17 View all Activated ACL 697
27 802.1X AUTHENTICATION CONFIGURATION 699
27.1 OVERVIEW 699
27.2 PRINCIPLES 699
27.2.1 Basic Concepts of 802.1X 700
27.2.2 Authentication Trigger Mode of 802.1X 701
27.2.3 Authentication Method of 802.1X 701
27.2.4 MAC Bypass Authentication 705
27.2.5 802.1X Authentication Supports Dynamic VLAN Authorization 705
27.2.6 802.1X Rapid Deployment 706
27.2.7 User Group Authorization Function 707
27.3 CONFIGURE 802.1X AUTHENTICATION 707
27.3.1 Global802.1X Authentication Enablement 707
27.3.2 802.1X Authentication Port Authorization Mode 708
27.3.3 802.1X Authentication Port Controlled Direction 708
27.3.4 802.1X Authentication EAPOL Protocol Version 709
27.3.5 802.1X Authentication Port Silent Time 710
27.3.6 802.1x Authorization Port Reauthentication Interval 710
27.3.7 802.1X Authorization Server Timeout Time 711
27.3.8 802.1X Authorization Client Timeout Time 712
27.3.9 802.1X Authorization Message Retransmission Interval 712
27.3.10 802.1X Authorization Message Retransmission Interval 713
27.3.11 802.1x Authorization Port Reauthentication Mode 714
27.3.12 802.1X Authentication Port Initialization 715
27.3.13 802.1X Authorization Key Encryption Function 715
27.3.14 Display 802.1X Authentication Global Information 716
27.3.15 Display 802.1X Authentication Detailed Information 716
27.3.16 Display 802.1X Authentication Port Information 717
27.3.17 Display 802.1X Authentication Port Diagnosis Information 718
27.3.18 Display 802.1X Authentication Port Session Information 719

27.3.19Display 802.1X Authentication Port Message Statistics 720
27.3.20RADIUS Server Regeneration Interval 721
27.3.21RADIUS Server 721
28 ALARM CONFIGURATION 723
28.1 ENABLE PORT ALARM 723
28.2 DISABLE PORT ALARM 724
28.3 ENABLE POWER ALARM 724
28.4 POWER OFF WARNING 725
29 IPDT CONFIGURATION 726
29.1 OVERVIEW 726
29.2 CONFIGURE IPDT 726
29.2.1 Create Session and Enter the Session View 726
29.2.2 Configure Source IP 727
29.2.3 Configure Destination IP 728
29.2.4 Configure the Number of Per Detection 728
29.2.5 Configure Time Interval 729
29.2.6 Enable/Disable IPDT Session 730
29.2.7 Display IPDT Session 730
30 RMON CONFIGURATION 732
30.1 OVERVIEW 732
30.2 PRINCIPLES 733
30.3 CONFIGURE RMON 736
30.3.1 RMON Alarm Group 736
30.3.2 RMON Statistical Group 738
30.3.3 RMON History Group 738
30.3.4 RMON Event Group 739
30.3.5 Display RMON Alarm Group Information 740
30.3.6 Display RMON Statistics Information 741
30.3.7 Display RMON History Group Information 741
30.3.8 Display RMON Event Group Information 742
31 LOG CONFIGURATION 743
31.1 LOG FILE SIZE LIMIT 743
31.2 LOG STDOUT DISPLAY 744
31.3 LOGINFORMATION HIGHEST DISPLAY LEVEL 744
31.4 LOG LEVEL RECORD DISPLAY 745
31.5 SYSLOG SERVER DOWNLOAD LOG 746
32 NTP CONFIGURATION 748
32.1 NTP SERVER 748
33 RTC CONFIGURATION 750
33.1 RTC ENABLE 750
33.2 DISPLAY RTC STATUS 751
34 NETWORK DIAGNOSE CONFIGURATION 752
34.1 PING TEST 752

34.2 TRACEROUTE TEST 753
34.3 PORT LOOPBACK 754
35 SYSTEM MAINTENANCE 755
35.1 DEVICE INFORMATION DISPLAY 755
35.1.1 Display System Version 755
35.1.2 Display Product Information 755
35.2 SYSTEM SOFTWARE UPGRADE 756
35.3 CONFIGURATION FILE IMPORT AND EXPORT 757
35.3.1 Import Configuration File 757
35.3.2 Configure File Export 757
35.4 LOG FILE EXPORT 758
35.5 SAVE CONFIGURATION 759
35.6 REBOOT THE DEVICE 760
35.7 RESTORE FACTORY SETTINGS 760

1 Login the Switch Configuration

1.1 Login the Switch Function Overview

There are two ways for users to manage devices: CLI and WEB.

- CLI
  After logging into the device through the Console port, Telnet, or SSH, use the command line provided by the device to manage and configure the device. This approach requires configuring the user interface for the corresponding login mode.
- WEB
  When the device acts as a server, users can log in to the device through the WEB administration. The device provides a graphical interface with the built-in WEB server to facilitate users to manage and maintain the device intuitively and conveniently. This method can only realize the management and maintenance part of functions of the device. If more complex or fine management of the device is needed, the CLI method is still needed.

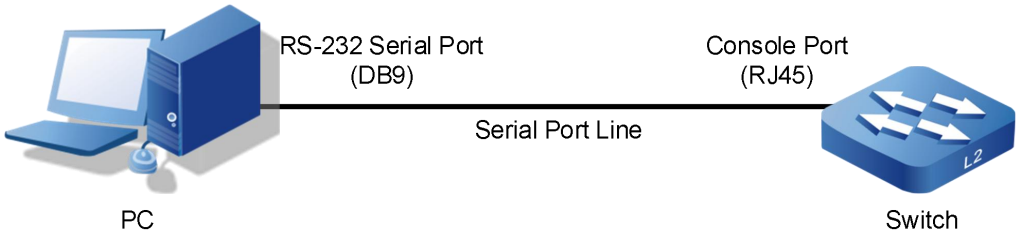
1.2 Login to the Switch

1.2.1 Login to the Switch via Serial Port

Logging in through the Console port is the basic way to log in a device, and is the basis for configuring a device logged in through other means. By default, users can log into the device directly through the serial port, and the switch baud rate is 115200bit/s. The PC can log into the command line interface of the device by connecting to its Console port.

Operation steps

Step 1 Connect the serial port of the computer to the Console port of the device through the serial port line to establish a local configuration environment, as shown in the topology diagram below.

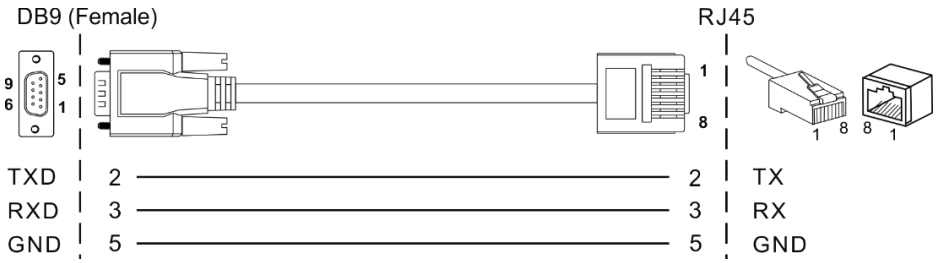


1 Connect DB9 at one end of serial port line to RS-232 serial port of PC.

2 Connect the RJ45 on the other end of the serial line to the Console port of the device.

Note:

Diagram of internal connection line of serial port line/communication cable is shown below.

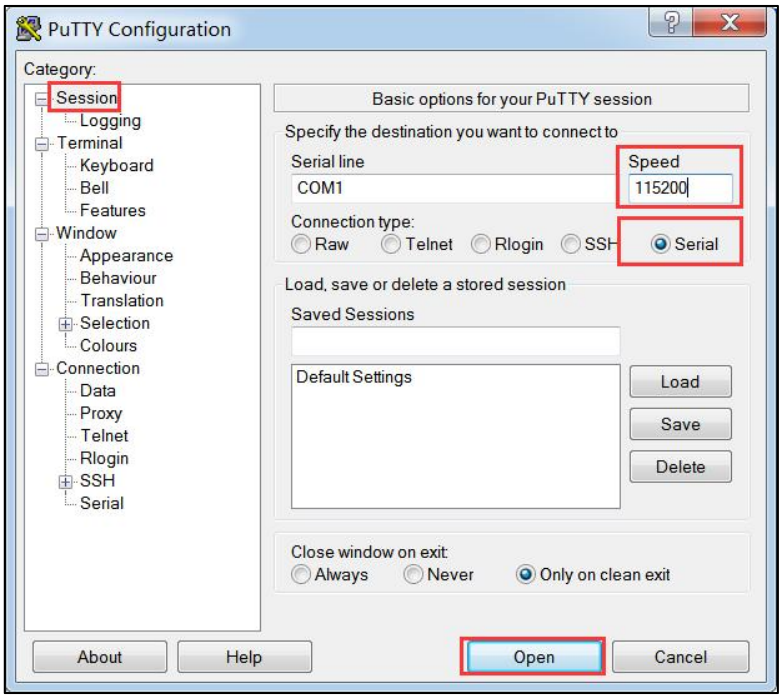


Step 2 Open the terminal simulation software on the PC, create a new connection, and set the interface and communication parameters of the connection. (Using PuTTY as an example here.)

1 Open PuTTY and click “Session” on the menu bar.

2 In the "Basic options for your PuTTY session" input box on the right, do the following:

- Select "Connection type" to "Serial".
- Enter "115200" in the "Speed" text box;
- Click “Open”.



3 The "COM1-PutTTY" command line edit dialog box pops up. Press enter key to enter user name and password. The user name and password is “admin123”, as shown in the following figure.

Step 3 End.

1.2.2 Login the Switch via Telnet

Log into the switch by Telnet, and the device acts as Telnet-Server By default, the Telnet-Server function is enabled. Therefore, before using Telnet to log into the switch, it is necessary to configure the IP of the switch through serial port to ensure the normal communication between PC and DUT.

Telnet-Server Configuration

Operation	Command	Remark
Enter Configure Mode	**switch# configure terminal**	-
Enable Telnet Server	**telnet-server enable**	Optional

Disable Telnet Server	`no telnet-server enable`	Optional

Notice

DUT acts as a Telnet server, if the logged client does not do any operation for a long time, it will automatically disconnect, i.e. timeout exit, and the function is enabled by default with 30m.

Through Telnet client login to the command line interface of the device, the client and the device should meet the following requires:

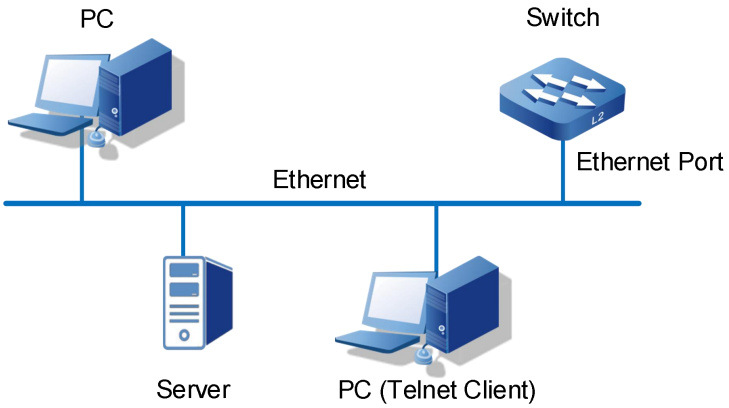
1 Configure the IP address of the switch correctly.

2 If the Telnet client and the device are in the same LAN, the IP address of the device and the client must be configured in the same network segment. Otherwise, the route between Telnet client and device must be accessible.

User can log in to the switch device through the Telnet client and configure the device if the two requires above are met.

Operation steps

Step 1 As shown in the figure below, set up the configuration environment to connect the Ethernet port of the computer to the Ethernet port of the device through the LAN.

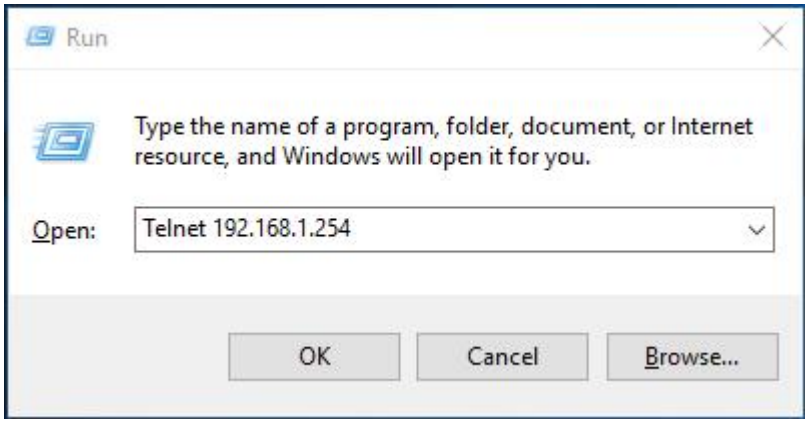


1 Run the Telnet client on the computer and input the administrative IP address of the Ethernet port connected the computer to the switch, as shown in the figure below.

2 Press "Win+R" to pop up the running window;

3 Enter “Telnet+ space + device IP address” in the “Open (O)” input box.

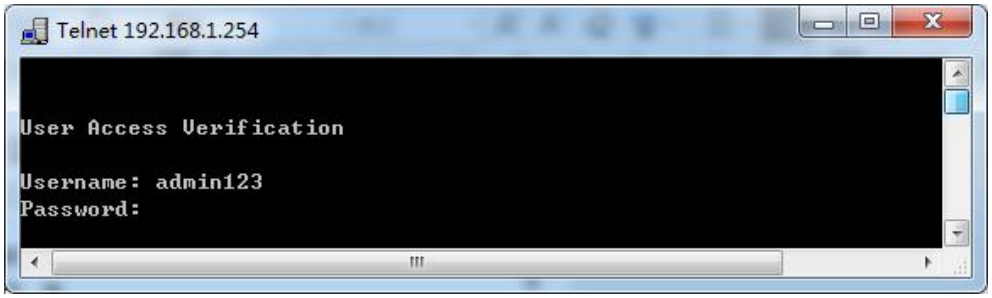
4 Click "OK” button.



Note:

- Using the command line prompt interface of Win7/Win8/Win10 and other operating systems to configure the device needs to enable Telnet client in advance, user can check and enable Telnet client in the Windows function window under the path of "Control Panel > Program and Function > Enable or Disable Windows function", if Telnet client has been enabled, user can ignore this instruction.
- If the computer operating system does not support Telnet clients, a third party software PuTTY can be used as a Telnet client.
- The default IP address of the device is “192.168.1.254”.

Step 2 The "Telnet" dialog box pops up and user can enter user name and password according to the hint. The user name and password is “admin123”, as shown in the following figure.



Step 3 End.

1.2.3 Login to the Switch via SSH

The switch can be used as an SSH server, but can not be used as an SSH client.

By default, the SSH server function of the device is disabled. Therefore, before using SSH to log in to the device, it is necessary to log in to the device through the Console port first, and enable the SSH server function and other properties of the device for corresponding configuration, so as to ensure normal login to the device through SSH.

SSH Configuration

Operation	Command	Remark

Operation	Command	Remark
Enter Configure Mode	switch# configure terminal	-
Enable SSH server	ssh-server enable	Optional
Disable SSH server	no ssh-server enable	Optional

Notice

If SSH login to DUT is needed, the simplest operations are:

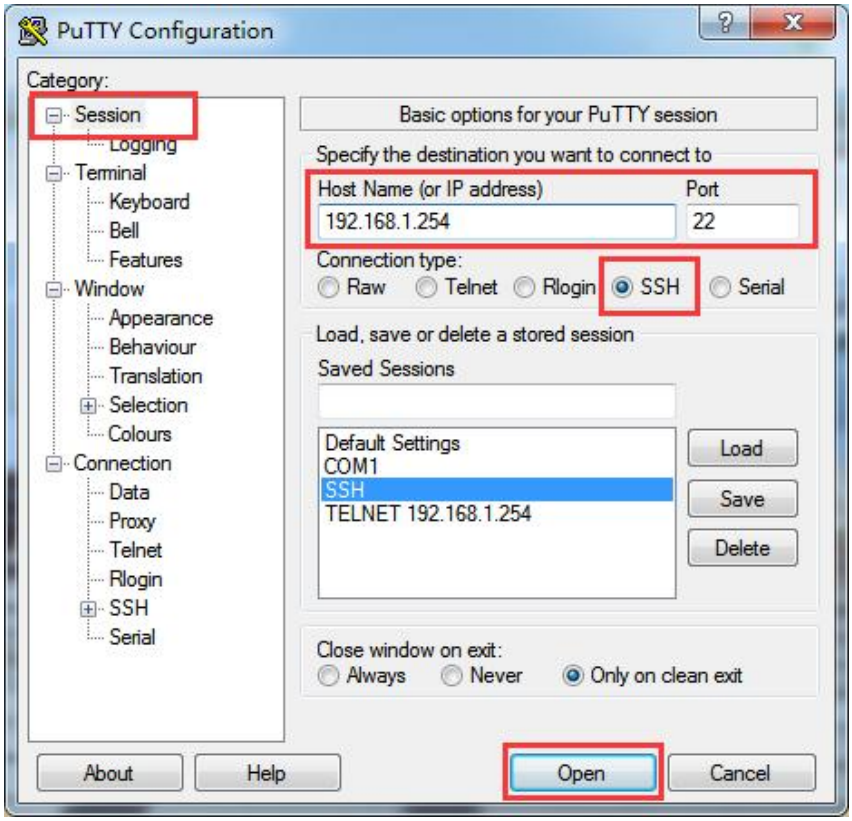
- Enable SSH;
- Configure SSH users, that is device users;
- Login the device.

Operation steps

Step 1 Using the Console port, enable SSH service using the "ssh-server enable" command.

```
switch> enable
switch# configure terminal
switch(config)# ssh-server enable
```

Step 2 Run the third-party PuTTY software on the PC host as an SSH client to establish a secure connection with the device, and fill in the following parameters:



1 Click “Session” in the “Category” bar;

2 Choose “SSH” in the “Connection type”;

3 Enter the IP address "192.168.1.254" of the device in the "Host Name (or IP address)" text box.

4 “Port” port number defaults to 22.

5 (Optional) enter the session name in the "Saved Sessions", such as SSH; click "Save" to save the session;

6 Click "Open" button to enter the SSH configuration interface;

7 Enter the user account name of this device in the SSH client, such as the default user name and password is "admin123";

Note:

With SSH enabled, all users on the device support SSH encrypted login.

8 Access to the device through SSH is successful, end.

Step 3 End.

1.2.4 Login the Switch via WEB:

User can log into the switch through the WEB, but the functions on the WEB are not complete and most functions cannot be configured on the WEB. It is not recommended to manage the switch in this way.

By default, the function that switch acts as an HTTP server is enabled. Before logging into the WEB, user needs to ensure that the client has a browser and corresponding IP address to ensure normal communication between the client and the HTTP server.

WEB Log in Configuration

Operation	Command	Remark
Enter global mode	**switch# configure terminal**	Required
Enable HTTP server	**http-server enable**	Optional
Disable HTTP server	**no http-server enable**	Optional

Configuration Environment Requirements

Client requirements: IE browser 8.0 above, some versions of 360 browser may have problems, other browsers have not found any problem at present.

Server requirements: switch configuration:

```
*Switch>enable
```

```
*Switch#configure terminal
# Switch enable WEB service
*Switch(config)#http-server enable
```

Login WEB Management Platform

The user enter http://X.X.X.X directly in the browser (default switch management IP is 192.168.1.254), press Enter key to enter the switch login interface, enter user name and password and click login to enter the main interface. The default user name and password of the device is “admin123”.

1.2.5 Manage the switch via Network Management Software

The switch supports login management via network management software. By default, SNMP function is enabled, and can use the default community name. This only shows that the switch can be managed by SNMP. Please refer to the SNMP user manual for more detailed configuration.

SNMP login configuration

Operation	Command	Remark
Enter global mode	**switch# configure terminal**	Required
Enable SNMP server	**snmp-server**	Required
Disable SNMP server	**no snmp-server**	Optional

1.3 Command Line

1.3.1 Command Analysis

Command consists of two parts: command word and command parameter. Commands are all lowercase, input is case-insensitive; command words come in many forms, including: capital letters, (), <>, *, etc.

For example: IP address A.B.C.D/M (secondary|), IP and address are command words, and A.B.C.D/M and (secondary|) are command parameters.

1.3.2 Command Line Mode

Here are four major command-line patterns:

- Exec Mode: also called "View Mode" , the basic mode of entering CLI. The prompt is ">", and user can only execute some simple commands, such as: show, enable, logout, etc
- Privileged Exec Mode: also known as "Enable Mode" with the prompt of "#", in Exec Mode, entered by executing enable command, or switched from other modes. Basic commands such as: debug, show, reboot, cp can be executed.
- Configure Mode: also known as "Configure Terminal", the prompt is "(config)#". User can execute the configure terminal to enter this mode in Privileged Exec mode, or switch to this mode from another mode, and all Configure Mode commands can be executed.
- Interface Mode: prompt is "(config-IFNAME)#". User can enter "Interface IFNAME" in ConfigurE Mode or switch to this Mode from another mode. Configuration command for the specified Interface can be executed.

1.3.3 Shortcut Key

Only the commonly used command parameters are covered here.

Shortcut Key	Note
?	Help command, enter"?" Command help is displayed.
Tab	Command completion, "Tab" can prompt or complete the remaining characters to be input when typing part of the command word.
Ctrl+D	To exit the current mode, can exit to the upper level mode in any mode, such as Interface Mode to Configure Mode.
Ctrl+C	End up command input or execution.
Ctrl+W	Delete an input command word or delete an input command parameter.
Ctrl+U	Deletes all characters from the current input command line.

【Instance】

"?" Help command. When using the command line, type“? ”command help is displayed. Cases are as follows:

1 Type only "? "in a configuration mode, a list of all commands in the current mode is displayed.

```
Switch#?
Exec commands:
```

```
  clear      Reset functions
  clock      Config clock time
  configure   Enter configuration mode
  copy       Copy file
  debug      Debugging functions (see also 'undebug')
  disable    Turn off privileged mode command
  dot1x      IEEE 802.1X Port-Based Access Control
  enable     Turn on privileged mode command
  erase      Erase file
  exit       End current mode and down to previous mode
  faults     Fault management command
  help       Description of the interactive help system
  logout     Exit from the EXEC
  loopback    config l2 interface loopback
  mstat     Show statistics after multiple multicast traceroutes
  mtrace     Trace multicast path from source to destination
  no         Negate a command or set its defaults
  ping       Send echo messages
  quit       Exit current mode and down to previous mode
  reboot     Halt and perform a cold restart
  reload     Halt and perform a cold restart
  rm         erase file
  rmon       Debugging functions (see also 'undebug')
  show       Show running system information
  ssh        Open a SSH connection
  telnet     Open a telnet connection
  terminal    Set terminal line parameters
  traceroute  Trace route to destination
  undebug    Disable debugging functions (see also 'debug')
  write        Write running configuration to memory, file or
terminal
```

2 All commands matching the current command word are displayed when partial command words are entered.

```
Switch#c?
  clear     Reset functions
  clock     Config clock time
  configure  Enter configuration mode
  copy      Copy file
```

1.4 Common Command

1.4.1 Password Verification

The configuration process for enabling password verification:

```
Switch>enable
Switch#configure terminal
Switch(config)#username admin password admin
Switch(config)#line vty 0
Switch(config-line)#login local
```

1.4.2 Customization Display

Currently, there are two ways: exclude and include:

- Exclude: only shows rows that do not contain the current string;
- Include: only displays the line that contains the current string;

Instance 1: show interface brief all

```
Switch#show interface brief all
Interface        IP-Address       Link     Protocol
lo               127.0.0.1/8       up       up
vlanif1          192.168.1.254/24   up       up

Interface        Link     Speed    InUti    OutUti   Duplex  Type
PVID Description
ge1              down     auto     0.000%   0.000%   auto    access  1
ge2              up       100m(a)  0.001%   0.001%   full(a) trunk   1
ge3              down     auto     0.000%   0.000%   auto    access  3
ge4              down     auto     0.000%   0.000%   auto    access  4
ge5              down     auto     0.000%   0.000%   auto    access  1
ge6              down     auto     0.000%   0.000%   auto    access  1
ge7              down     auto     0.000%   0.000%   auto    access  1
ge8              down     auto     0.000%   0.000%   auto    access  1
ge9              down     auto     0.000%   0.000%   auto    access  1
ge10             down     auto     0.000%   0.000%   auto    access  1
ge11             down     auto     0.000%   0.000%   auto    access  1
ge12             down     auto     0.000%   0.000%   auto    access  1
ge13             down     auto     0.000%   0.000%   auto    access  1
ge14             down     auto     0.000%   0.000%   auto    access  1
ge15             down     auto     0.000%   0.000%   auto    access  1
ge16             down     auto     0.000%   0.000%   auto    access  1
```

```
xe1          down     10g(a)   0.000%   0.000%  full(a) access  1
xe2          down     10g(a)   0.000%   0.000%  full(a) access  1
xe3          down     10g(a)   0.000%   0.000%  full(a) access  1
xe4          down     10g(a)   0.000%   0.000%  full(a) access  1
```

Instance 2: show interface brief all | exclude ge Does not show interfaces containing ge

```
Switch#show interface brief all | exclude ge
Interface       IP-Address      Link     Protocol
lo             127.0.0.1       up       up
vlanif1         192.168.1.254    up       up

Interface       Link     Speed   Duplex  Type    PVID Description
xe1            down     10g(a)  full(a) access  1
xe2            down     10g(a)  full(a) access  1
xe3            down     10g(a)  full(a) access  1
xe4            down     10g(a)  full(a) access  1
```

Instance 3: show interface brief all | include ge Only show interfaces containing ge

```
Switch#show interface brief all | include ge
ge1            down     auto    auto    access  1
ge2            down     auto    auto    access  1
ge3            down     auto    auto    access  1
ge4            down     auto    auto    access  1
ge5            down     auto    auto    access  1
ge6            down     auto    auto    access  1
ge7            down     auto    auto    access  1
ge8            down     auto    auto    access  1
ge9            down     auto    auto    access  1
ge10           down     auto    auto    access  1
ge11           down     auto    auto    access  1
ge12           down     auto    auto    access  1
ge13           down     auto    auto    access  1
ge14           down     auto    auto    access  1
ge15           down     auto    auto    access  1
ge16           down     auto    auto    access  1
```

1.4.3 Configuration Management

Command	Note
Switch#**show running-config**	Displays the configuration of the current system running
Switch#**show startup-config**	Displays the configuration of

Command	Note
	the system startup profile
Switch#**write**	Save command
Switch#**erase startup-config**	Restore factory settings
Switch#**copy tftp startup-config** 192.168.1.168 **Switch.conf**	Upload configuration file to switch
Switch#**copy flash startup-config** 192.168.1.168 **Switch.conf**	Download Configuration file from switch

Notice

If the configuration of the current system is inconsistent with the configuration of the system startup configuration file:

- After entering Configure Mode, the prompt is "*Switch";
- Performing a system reboot will prompt whether need to save a disk.

1.4.4 System Upgrading

Suppose the upgrade package is "packetapp.bin", the IP address of the TFTP server is "192.168.1.168", and the command to upgrade the switch system is:

```
Switch#copy tftp package 192.168.1.168 packetapp.bin
```

Notes

The system upgrade must restart the switch to take effect.

The command to restart the system is: Switch#reboot.

After executing the restart, the screen will display as follows:

```
*Switch#reboot
save running config? (y/n): y
Building configuration...
[OK]
reboot system? (y/n): y
```

1.4.5 Debug Mode

Enter the debugging state of function module.

```
*Switch#debug ospf packet /*Enable receiving packet debug of ospf*/
*Switch(config)#log stdout /*Log output to serial port*/
```

2 User Configuration

2.1 Add User

【Command】

```
username STRING
username STRING login-control closet <0-900>
username STRING password PASSWD (debugger|ssh-user|)
username STRING password-control aging <0-365> <30-86399>
username STRING privilege <0-15>
username STRING privilege <0-15> password PASSWD (ssh-user|)
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

STRING: user name, which can only contain numbers, characters and "!", "@", "_", "-",etc. The default user priority is 15.

<0-900>: locking time of secure login, unit: seconds.

PASSWD: password, including numbers, characters, "~", "!", “@”, “#”, “$”, “%”, “_”, “-“, ”.”.

<0-365> <30-86399 >: password aging time, < 0-365 > unit: days, < 30-86399 > unit: seconds.

<0-15>: a total of 16 user privilege priorities, divided into four categories:

- 0: visit level; user can only view device version information and some simple configuration information.
- 1: view level; The configuration information of the device can be viewed, but the configuration of the device cannot be modified.

- 2: configuration level; User can view the configuration information of the device and configure some functional parameters of the device, but cannot manage the device.
- 3-15: manage level, user has all privileges of the device, including downloading, uploading, rebooting, modifying device information and other other operations.

【Description】

username STRING : this command adds a user name that has no password and privilege defaults to 15.

username STRING **login-control closet** <0-900>: indicates the user login locking interval. According to the security level, if a user fails to verify more than 5 times in 10 minutes, it will be locked for a specified time, and the default value is 600s. During this locking period, if verification fails agian, the locking time will be updated.

username STRING **password** PASSWD **(debugger|ssh-user|)**: if the user is forced to be a debugger user, the user can access the device through Telnet 2323 port, which is convenient for debugging. If the user is forced to be a ssh-user, the user can log in to the device through SSH. By default, all newly created users do not support SSH login.

username STRING **password-control aging** <0-365> <30-86399>: configure the aging time of user password, which is 90 days by default. After password aging, you will be prompted to change the password. Parameter 1 is days, and parameter 2 is seconds.

username STRING **privilege** <0-15>: this command can set privileges for users.

username STRING **privilege** <0-15> **password** PASSWD **(ssh-user|)**: this command can create a new user, specify privilege, password and user type.

【Instance】

```
Switch> enable
Switch#configure terminal
//create user admin123
Switch(config)#username admin123 privilege 15 password admin123
//set the aging time of admin123's password to 10 days and 1 hour
switch(config)# username admin123 password-control aging 10 3600
// Set the login failure locking time of user admin123 to 100 minutes
switch(config)# username admin123 login-control closet 100
```

2.2 Delete User

【Command】

no username STRING

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

STRING : user name, which can only contain numbers, characters, “!”, “@”, “_”, “-“, etc..

【Description】

Delete specified user.

【Instance】

Delete user admin123.

```
Switch> enable
Switch#configure terminal
Switch(config)#no username admin123
```

Notice

Deleted user names cannot be logged into the device, and when all user names have been deleted, the device only can be logged in through the Console port.

2.3 View Current Online Users

【Command】

show users

show logined users

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show logined users: view current online users

show users: view current users with lower priority.

【Instance】

```
Switch> enable
Switch#show Logined users
Line     User     Type    Idle     Host(s)    Uptimes      Location
 0     admin123  console   0                  01:36:08     console
```

2.4 Console Login Management

The Console user interface is used to manage and monitor users logging in through the Console port. The device provides a RJ45 type Console port of RS-232 serial port. The terminal serial port of the user can connect directly with the device Console port to achieve local access to the device.

【Command】

line console <0-0>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<0-0> : parameter "0", local Console configuration, supporting only one user.

【Description】

Line console <0-0>: enters the Console user interface configuration view. The Console user interface can configure connection timeout, password validation, priority, and history command buffer sizes.

【Instance】

```
*Switch#configure terminal
*Switch(config)#line console 0
*Switch(config-console_0)#
```

2.5 Virtual Terminal Login Management

The VTY (Virtual Type Terminal) user interface is used to manage and monitor users logging in through VTY. After the user establishes a Telnet or SSH connection with the device through the terminal, a VTY channel is established. Currently each device supports up to 16 simultaneous VTY users. There is no fixed relationship between user interface and user. The user interface is assigned differently when the same user logs in different ways. Different user interfaces may be assigned for different login times for the same user.

【Command】

line vty <0-15>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<0-15> : VTY user channel 0-15, supports 16 VTY users to access the device simultaneously.

【Description】

line vty <0-15>: enters the VTY user interface configuration view. The VTY user interface can configure connection timeout, password validation, priority, and history command buffer sizes.

【Instance】

```
*Switch#configure terminal
*Switch(config)#line vty 0
*Switch(config-vty_0)#
```

2.6 Timeout Logout

【Command】

exec-timeout <0-35791> <10-2147483>

【View】

Console/ VTY user interface configuration mode

【Default Level】

2: Configuration level

【Parameter】

<0-35791> : timeout minutes range.

<10-2147483> : timeout seconds range.

【Description】

exec-timeout <0-35791> <10-2147483> : this command disconnects idle connections within a set time. If the connection is always idle during the set time, the system will automatically disconnect the connection. By default, the timeout of user interface disconnection is 10 minutes.

【Instance】

The system is configured with a 10-minute timeout by default. If the user is configured with password authentication, the user needs to enter the username and password again after the timeout to enter the system.

Configuration process for modifying the timeout logout:

```
Switch>enable
Switch#configure terminal
Switch(config)#line vty 0
Switch(config-vty_0)#exec-timeout 0 600
```

3 Port Configuration

3.1 Enter Port Configuration Mode

【Command】

```
interface IFNAME
interface ge <1-24>
interface loopback <0-1>
interface po <1-12>
interface range (ge | xe)
interface sa <1-12>
interface vlanif <1-4094>
interface xe <1-4>
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

IFNAME: port name

ge: Gigabit port name.

loopback: loopback port name

po: dynamic aggregation group name.

range: supports range type port input. For example, interface range ge 1-10 is denoted as going into Gigabit port 1-10. Only Gigabit ports and 10 Gigabit ports are currently supported.

sa: static aggregation group name

vlanif : layer 3 interface

xe: 10 Gigabit port.

【Description】

This command is the mode navigation command that goes from Configure Mode to interface configuration mode. The next step is to modify the configuration of the corresponding interface.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface range ge 1-10
Switch(config-ge1-10)#
Enter 10 ports from ge1 to ge10.
```

3.2 Port Rate Limit

【Command】

bandwidth <64-10000000>
no bandwidth

【View】

fe (100M Ethernet) port view
ge (Gigabit Ethernet) port view
xe (10 Gigabit Ethernet) port view
sa (static aggregation group) port view
po (dynamic aggregation group) port view

【Default Level】

2: Configuration level

【Parameter】

<64-10000000> : the unit is kbps. For different ports, there are some restrictions on the parameters. The allowed input range of normal Gigabit ports is 64-1000000, and the allowed input range of 10 Gigabit ports is 64-10000000. If the input parameter is not in the specified range, and the input parameter is not a multiple of 64, setting will not be successful and an error will be returned.

【Description】

bandwidth: this command does not actually affect the bandwidth of an interface, but simply allows the user to inform the system the bandwidth standard of that interface. By default, the bandwidth of an Ethernet interface is determined by the rate of the actual port connection, and can be manually configured if necessary. The bandwidth

is only a routing parameter and does not affect the real bandwidth of the interface of the physical link.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#bandwidth 128
```

3.3 Port Settings

3.3.1 Combo PortTransmission Media

【Command】

combo-port (auto | copper | fiber)

【View】

fe (100M Ethernet) port view

ge (Gigabit Ethernet) port view

【Default Level】

2: Configuration level

【Parameter】

auto：Automatically select media types

Copper：forced to select copper port.

fiber：forced to select fiber port.

【Description】

combo-port: used to select the media type of the port.

Notes

- Currently only the first eight ports can set the media type of the port.
- Change the media type of the port, the properties of the port will be updated to the default values.

By default, the first eight ports (ge1-ge8) default to "auto" mode, and the fiber transmission medium is preferred. That is, if a port is connected to both fiber port and copper port, then the port type is fiber port. If a port is only connected a copper port, the port type is copper port.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#combo-port copper
```

3.3.2 Duplex Mode

【Command】

duplex (auto | full | half)

no duplex

【View】

fe (100M Ethernet) port view

ge (Gigabit Ethernet) port view

xe (10 Gigabit Ethernet) port view

【Default Level】

2: Configuration level

【Parameter】

Auto: full duplex and half duplex self-adaption.

full: represents full duplex.

half: represents half duplex.

【Description】

duplex (auto | full | half) : this command is used to set the duplex mode of the port.

By default, duplex mode of all ports is auto.

When setting the normal port rate to Gigabit, the duplex mode of the port cannot be set to half-duplex. When setting 10 Gigabit fiber ports, the duplex mode of the port cannot be set to half duplex. Otherwise the setting will not take effect and an error will be returned.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#duplex full
```

3.3.3 Flow Control

【Command】

```
flowcontrol (both| receive | send)
flowcontrol send (on | off)
flowcontrol receive (on | off)
no flowcontrol
```

【View】

fe (100M Ethernet) port view

ge (Gigabit Ethernet) port view

xe (10 Gigabit Ethernet) port view

sa (static aggregation group) port view

po (dynamic aggregation group) port view

【Default Level】

2: Configuration level

【Parameter】

both: Data transmit and receive of the port are set to self-negotiate flow control.

receive (on | off) : only enable or disable flow control on data receiving of the ports.

send (on | off) : only enable or disable flow control on data transmission of the ports.

【Description】

flowcontrol: this command is used to enable or disable flow control of the ports.

By default, flow control on all ports is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#flowcontrol both
```

3.3.4 Max-Frame

【Command】

```
mtu <64-16360>
no mtu
```

【View】

fe (100M Ethernet) port view

ge (Gigabit Ethernet) port view

xe (10 Gigabit Ethernet) port view

sa (static aggregation group) port view

po (dynamic aggregation group) port view

vlanif (layer 3) port view

【Default Level】

2: Configuration level

【Parameter】

<64-16360> : the allowed setting range of mtu is 64-16360.

<128-1500> : the allowed setting range of mtu in a layer 3 interface is 128-1500.

【Description】

mtu: this command is used to set the maximum data frame length supported by the interface, that is, the maximum length of the data portion of the link.

By default, the maximum data frame length for all physical ports is set to 1518. The MTU for the virtual port, such as vlanif1, is set to 1500.

Notice

When setting up virtual ports such as vlanif1, the maximum MTU value allowed is 1500.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#mtu 1800
```

3.3.5 Interface Switch

【Command】

shutdown

no shutdown

【View】

fe (100M Ethernet) port view

ge (Gigabit Ethernet) port view

xe (10 Gigabit Ethernet) port view

sa (static aggregation group) port view

po (dynamic aggregation group) port view

vlanif (layer 3) port view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

shutdown: for interfaces (Ethernet ports, converged ports, and switched virtual interfaces), the command is primarily to close the corresponding interface, but other configurations of the interface still exist, just do not work. no shutdown is to open the port.

By default, the administrative state of the interface is UP.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface sa1
Switch(config-sa1)#shutdown
```

3.3.6 Rate

【Command】

```
speed (auto | 10m | 100m | 1g | 10g)
speed (auto | 1g | 10g)
speed (auto | 10m | 100m | 1g )
no speed
```

【View】

ge (Gigabit Ethernet) port view

xe (10 Gigabit Ethernet) port view

sa (static aggregation group) port view

po (dynamic aggregation group) port view

【Default Level】

2: Configuration level

【Parameter】

auto: Indicates that the rate of the interface is self-adaptive.

10m: set the interface rate to 10Mbps.

100m: set the interface rate to 100Mbps.

1g: set the interface rate to 1Gbps.

10g: set the interface rate to 10Gbps.

(auto|10m|100m|1g|10g): port rate configuration of dynamic and static aggregation groups.

(auto|1g|10g): 10 Gigabit port speed configuration

(Auto|10m|100m|1g): Gigabit port speed configuration.

【Description】

speed: this command is used to set the rate of the port.

By default, the rate of the interface is self-adaptive (auto). The port rate cannot be set to 1g or above when setting the normal port duplex mode to half. The port rate is set to a minimum of 1g, when setting a 10 Gigabit fiber port.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#speed 100m
```

3.4 Port Isolation

【Command】

port-isolate enable group <1-8>

no port-isolate enable

【View】

ge (Gigabit Ethernet) port view

xe (10 Gigabit Ethernet) port view

sa (static aggregation group) port view

po (dynamic aggregation group) port view

【Default Level】

2: Configuration level

【Parameter】

<1-8> : isolation group ID

【Description】

port-isolate: this command is used to add the current Ethernet ports to the isolation group.

no port-isolate : This command is used to remove the current Ethernet ports from the isolation group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#port-isolate enable group 1
Switch(config-ge8)#no port-isolate enable
```

3.5 Storm Suppression

【Command】

storm-control (broadcast | dlf | multicast) level LEVEL
no storm-control (broadcast | dlf | multicast) level

【View】

ge (Gigabit Ethernet) port view

xe (10 Gigabit Ethernet) port view

sa (static aggregation group) port view

po (dynamic aggregation group) port view

【Default Level】

2: Configuration level

【Parameter】

broadcast: sets the limit of broadcast message traffic of the port

dlf: Destination look-up fail, which is to set unicast storm suppression.

multicast: sets the limit of multicast message traffic of the port

LEVEL: the percentage of restricted storm suppression, ranging from 0.00 to 100.00 to two decimal places.

【Description】

storm-control: this command is used to set the limit on unicast/multicast/broadcast messages traffic of the port.

no storm-control: this command is used to unconfigure restricted port messages. After setting the upper limit of port message traffic, the port regularly detects the received unicast/multicast/broadcast message flow. Once the data flow of a certain type of message is detected to reach the storm control of the port, it would be considered as storm, then the port can block the forwarding of such message.

By default, the percentage of storm suppression is 100.00%.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#storm-control broadcast level 20.02
Switch(config-ge8)#no storm-control broadcast level
```

3.6 MAC Address

3.6.1 Clear Dynamic MAC address

【Command】

clear mac-address-table dynamic (MAC | address MACADDR| interface IFNAME | vlan VID)

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

MAC: means clear of the specified dynamic MAC address, in the format HHHH.HHHH.HHHH.

address: means clear the specified dynamic MAC address.

interface: means to clear all dynamic addresses of a specified interface.

vlan: means to clear all dynamic addresses of the specified vlan, ranging from 1-4094;

【Description】

clear mac-address-table dynamic: this command is used to clear the specified dynamic MAC address, or to clear all dynamic MAC addresses on the specified interface or VLAN.

【Instance】

```
Switch> enable
*Switch#clear mac-address-table dynamic interface ge1
```

3.6.2MAC Address Learning

【Command】

mac-address-learning (disable | enable)

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

disable: disable global MAC learning.

enable: enable global MAC learning.

【Description】

mac-address-learning: The function of this command is to disable the global MAC address learning ability, so that the global MAC address learning can not be carried out; Or enable the global MAC address learning ability, according to the port of the MAC address learning ability to take effect.

By default, the learning capability of the global MAC address is enabled.

When the MAC address learning ability is enabled, the MAC address learned by the port is a dynamic MAC address, and the aging time is determined by the user settings.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mac-address-learning disable
```

3.6.3MAC Address Aging-Time

【Command】

mac-address-ageingtime (0 | <10-1000000>)

no mac-address-ageingtime

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

ageing-time: set the aging time of the global MAC. The value range of dynamic address aging time is <10-1000000>, in seconds. 0 means to disable aging function.

【Description】

mac-address-ageingtime: command is to set the dynamic aging time of the MAC address table. When the user enters the 0 parameter, it means to disable the aging time of MAC.

no mac-address-ageingtime: MAC address aging time is restored to the default value.

By default, the aging time is set to 300 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)##mac-address-ageingtime 500
```

3.6.4 Static MAC Address Filtering

【Command】

mac-address-table static MAC (discard | forward) IFNAME vlan VLAN
no mac-address-table static address MAC (vlan VLAN |)

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

Discard: set a MAC address for the port .Discard the packet if the packet's source MAC address is the same as the set MAC address.

Forward: sets a MAC address of the port. Forward the packet if the source MAC address is consistent with the set MAC address.

vlan: specifies the vlan corresponding to the table entry, with a range of <2-4094>. If there is no input, the default is vlan 1.

【Description】

mac-address-table static MAC forward IFNAME: this command is to set a static MAC address to the MAC table entry on the specified port. Static addresses, as

opposed to dynamic ones, never aging and can only be manually configured and deleted. Static addresses will not lost even if the device is reset.

No static address is set by default.

Static addresses cannot be set to multicast addresses.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mac-address-table static 0.8.4 forward ge10 vlan 3
```

3.6.5 Multicast MAC Address Filtering

【Command】

mac-address-table multicast MAC (discard | forward) IFNAMELIST vlan <2-4094>

no mac-address-table multicast address MAC interface IFNAME vlan <2-4094>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

IFNAMELIST: multiple ports can be entered simultaneously. For example: ge1, ge2

vlan: specifies the vlan corresponding to the table entry, with a range of <2-4094>. If there is no input, the default is vlan 1.

【Description】

mac-address-table static MAC forward IFNAMELIST: this command is to set a static Multicast MAC address to the MAC table entry on the specified port. Static addresses, as opposed to dynamic protocol learned, never aging and can only be manually configured and deleted. Static addresses will not lost even if the device is reset.

no mac-address-table multicast: is used to remove static multicast table entries configured with command. The three commands of interface, mac and vlan can be randomly combined. That is, through the port to batch delete, can also delete a specific MAC specific VLAN under the specific port.

No static address is set by default.

Notice

Currently only single port deletions are allowed when deleting configuration. Multiple port combination deletion is not allowed.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mac-address-table multicast 0100.5e00.0001
forward ge10,ge11,ge12 vlan 3
```

3.6.6 Display MAC Address Table

【Command】

```
show mac-address-table
show mac-address-table ( multicast | dynamic | static | )
```

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

multicast：displays the table entry of multicast MAC address.

Dynamic: displays dynamic MAC address table entries.

static: displays dynamic MAC address table entries.

【Description】

show mac-address-table: this command displays the MAC table of the device. Without parameters, all MAC addresses are displayed, including user-configured static MAC addresses, dynamic MAC addresses learned by protocol, and multicast MAC addresses. With the relevant parameters, the corresponding MAC address is displayed, when with multicast, dynamic and static multicast MAC addresses are displayed.

【Instance】

```
Switch> enable
Switch#show mac-address-table
```

3.7 Mirror Command

3.7.1 Port Mirror Configuration

【Command】

mirror session <1-4> (both | receive | transmit) destination IFNAME source IFNAMELIST

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

session: mirror group number, value range: <1-4>.

TRAFFIC: messages direction of monitored port. That is to monitor the messages received or transmit.

- both: means both receiving and sending packets are monitored.
- transmit: stands for direction of transmit package.
- receive: stands for direction of receive package.

directions: means the destination port.

source: means the source port. Multiple ports can be input at the same time, separated by commas.

【Description】

Mirrors a message in the specified direction of the source port to the destination port.

Notice

There is and can only be one destination port for mirroring, but multiple source ports can be configured simultaneously. And the destination port of one mirror group cannot be the source port for other mirror groups.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mirror session 1 both destination ge1 source ge2,ge3
```

3.7.2 Delete Port Mirror

【Command】

no mirror session <1-4> direction (both | receive | transmit) source IFNAME

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

session: mirror group number, value range: <1-4>.

TRAFFIC: messages direction of monitored port. That is to monitor the messages received or transmit.

- both: means both receiving and sending packets are monitored.
- transmit: stands for direction of transmit package.
- receive: stands for direction of receive package.

source: means the source port. Multiple ports can be input at the same time, separated by commas.

【Description】

Delete Mirror Configuration The three parameters in this instruction are optional, that is, user can only type session, or session and direction when deleting, or directly type no mirror without any parameters. And parameter position arbitrary swap, will not affect the execution of instructions.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#no mirror session 1 source ge2
```

3.8 Link Aggregation Configuration

3.8.1 Dynamic Aggregation System Priority

【Command】

lacp system-priority <priority>

no lacp system-priority

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<priority> : dynamic aggregation system priority, range is 1-65535

【Description】

lacp system-priority: this command is used for dynamical aggregate system priorities.

By default, the system priority is 32768.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#lacp system-priority 1
```

3.8.2 Dynamic Aggregation Port Priority

【Command】

```
lacp prot-priority <priority>
no lacp prot-priority
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

<priority> : dynamic aggregation port priority, range is 1-65535.

【Description】

lacp port-priority: this command is used for dynamic aggregation port priorities.

By default, the port priority is 32768.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#lacp port-priority 1
```

3.8.3 Dynamic Aggregation Port Timeout

【Command】

lacp timeout (short | long)

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

Short: short timeout 3 seconds, the time threshold of neighborhood information aging.

Long: long timeout 90 seconds, the time threshold of neighborhood information aging.

【Description】

lacp timeout: this command is used for dynamic aggregate port timeout.

By default, it is long timeout.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#lacp timeout short
```

3.8.4 Add Dynamic Aggregation Group

【Command】

channel-group <id> mode (active | passive)

no channel-group

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

< id> : aggregation group number, the range is <1-12>.

Active: active mode, in which the switch actively initiates the aggregation negotiation process.

Passive: the mode in which the switch passively receives the aggregate negotiation process.

【Description】

channel-group: this command is used to add dynamic aggregation port members and configure the LACP mode for the ports.

When the first aggregation group member port is added, the corresponding aggregation group interface will be created. The interface name is Po + aggregation group number (the static aggregation group is sa+ aggregation group number). For example, a dynamic aggregation group interface named po100 with aggregation group number 100 is created by command channel-group 100 mode active.

When the last aggregation group member port is deleted, the corresponding aggregation group interface will be deleted.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#channel-group 1 mode active
```

3.8.5 Add static LACP

【Command】

static-channel-group <id>

no static-channel-group

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

< id> : aggregation group number, the range is <1-12>.

【Description】

static-channel-group: this command is used to add static aggregate port members.

When the first aggregation group member port is added, the corresponding aggregation group interface will be created. The interface name is sa + aggregation group number (the dynamic aggregation group is po+ aggregation group number). For

example, a static aggregation group interface named sa9 with aggregation group number 9 is created by command static-channel-group 9.

When the last aggregation group member port is deleted, the corresponding aggregation group interface will be deleted.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#static-channel-group 2
```

3.8.6 Link Aggregation Load Balance Mode

【Command】

port-channel load-balance (dst-ip | dst-mac | dst-port | src-dst-ip | src-dst-mac | src-dst-port | src-ip | src-mac | src-port)

no port-channel load-balance

【View】

Aggregation group interface view

【Default Level】

2: Configuration level

【Parameter】

dst-ip: Load balance mode based on destination IP;

src-ip: Load balance mode based on source IP;

src-dst-ip: Load balance mode based on source and destination IP;

dst-mac: Load balance mode based on destination MAC;

src-mac: Load balance mode based on source MAC.

src-dst-mac: Load balance mode based on source and destination MAC;

dst-port: the load balance mode is based on destination port, do not support currently.

src-port: the load balance mode is based on source port, do not support currently.

src-dst-port: the load balance mode is based on source and destination port, do not support currently.

【Description】

port-channel load-balance: this command is used to configure the load balance mode of the aggregate group.

【Instance】

```
Switch> enable
```

```
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#static-channel-group 2
Switch(config)#interface sa2
Switch(config-sa2)#port-channel load-balance dst-port
Switch(config-sa2)#exit
Switch(config)#interface ge7
Switch(config-ge7)#channel-group 1 mode active
Switch(config)#interface po1
Switch(config-po1)#port-channel load-balance src-mac
```

3.8.7 Displays Dynamic Aggregation Group

【Command】

show etherchannel {[<*id*>] | [detail] | [load-balance] | [summary]}

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

<id>: LACP aggregation group number

【Description】

show etherchannel: this command is used for LACP aggregation group related information.

【Instance】

```
Switch#show etherchannel
% Lacp Aggregator: po1
% Member:
   ge7
```

3.8.8 Displays Static Aggregation Group

【Command】

show static-channel-group

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show static-channel-group: this command is used for static aggregation group information.

【Instance】

```
Switch#show static-channel-group
% Static Aggregator: sa1
% Member:         state
   ge8            unbndl
```

3.9 Aggregation Protection

3.9.1 Enable Aggregation Protection

【Command】

static-channel-group detect
no static-channel-group detect

【View】

Sa Port view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

static-channel-group detect: enable link aggregation protection.
no static-channel-group detect: disable link aggregation protection.

【Instance】

```
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#static-channel-group 2
*Switch(config)#interface ge2
```

```
*Switch(config-ge2)#static-channel-group 2
*Switch(config)#interface sa2
*Switch(config-sa2)#static-channel-group detect
```

3.9.2 Displays Aggregation Protection Status Information

【Command】

```
show static-channel-group detect
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show static-channel-group detect: Displays link aggregation protection status.

【Instance】

```
Switch#show static-channel-group detect
  Interface sa2:
       LinkStatus      : Admin Enable, Link Up
       Default Vlan    : 1
       Neighbor        : 0000.0000.0000
       Role            : Slave
       Master Port     : None
       Error State     : None

       Interface ge1:
       LinkStatus      : Admin Enable, Link Up
       Detection Times : 0
       Priority        : 1
       Neighbor        : None
       Expire          : 4
       Main Channel    : No
       Error Packets   : 0
       Error Lever     : 0
       Link Quality    : High
       Process State   : Normal
```

```
Interface ge2:
LinkStatus      : Admin Enable, Link Down
Detection Times : 0
Priority        : 1
Neighbor        : None
Expire          : 4
Main Channel    : No
Error Packets   : 0
Error Lever     : 0
Link Quality    : Low
Process State   : Normal
```

3.10 Port Statistics

Port message statistics can be seen through the show interface IFNAME command. The following is the analysis of this instruction.

3.10.1 Display Port

【Command】

show interface IFNAME

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

IFNAME: port name

【Description】

The display mainly includes port name, port number, port medium, port property, port MAC address, MTU, bandwidth, configuration rate, duplex mode, running time and port message statistics. Here is an explanation of the key words that appear in the statistics:

- "Input" refers to the message statistics received by the port, i.e., "receive". "output" means the number of packets transmitted by the port, i.e., "transmit".
- "TYPE" refer to the classification of statistical message types.

- "Total" represents the statistics of all types of messages in the corresponding direction (i.e. input or output), and the unit is bit.
- "Unicast" represents the statistics of the packets of Unicast in the corresponding direction, and the unit is packets.
- "Multicast" represents the statistics of the packets of packets in the corresponding direction, and the unit is packets.
- "Broadcast" represents the statistics of the number of packets Broadcast in the corresponding direction, and the unit is packets.
- "Dropped" represents the statistics of the packets lost in the corresponding direction, and the unit is packets.
- "Error" represents the statistics of the number of packets of errors in the corresponding direction, with the unit of packets.

- "RATE" refers to the rate in the corresponding direction (it should be noted that this rate refers to the average rate in the corresponding type and direction in a specific period of time. Port statistics cannot be updated in real time, with an interval of about 24 seconds). In the "Total" type, the rate is expressed in parentheses as the ratio of message bytes to the "defined bandwidth" of the port, not the set bandwidth. The "defined bandwidth" here refers to the maximum bandwidth corresponding to the port, that is, the bandwidth ratio of the Gigabit port is calculated by Gigabit, and the 10 Gigabit port is calculated by 10 Gigabit, having nothing to do with the actual bandwidth set by the user.
- "PEAK" refers to the PEAK value, which is the maximum speed from the start to the execution of the command. The following brackets represent the time point at which the peak occurs.
- "TOTAL" refers to the total number of corresponding "TYPE". That is to say, the number (or digits) of corresponding type and direction messages obtained from startup to execution of command are counted. The following brackets represent the unit conversion result of the corresponding TOTAL (that is, when the TOTAL is 1024, the brackets show 1K).

In addition, all unit conversion in the command is carried out in the form of 1K = 1024.

【Instance】

```
Switch> enable
Switch#show interface ge1
```

3.11 Link Flapping Protection Configuration

3.11.1 Enable Link Flapping Protection

【Command】

link-flap protection enable

no link-flap protection enable

【View】

Port view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

The command is used to enable link flapping protection function.

By default, link flapping protection function is not enabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge2
Switch(config-ge2)#link-flap protection enable
```

3.11.2 Enable Link Flapping Auto-Recovery

【Command】

link-flap auto-recovery enable

no link-flap auto-recovery enable

【View】

Global configuration mode, port view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

Command is used to enable link flapping auto-recovery of global or port. If link flapping auto-recovery is enabled, after the interface enters the link flapping protection state, it will be delayed for a certain time (configured recovery interval), and the interface will exit the link flapping protection state.

By default, link flapping auto-recovery of global or port is not enabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge2
Switch(config-ge2)# link-flap auto-recovery enable
```

3.11.3 Configure Recovery Interval of Link Flapping

【Command】

```
link-flap auto-recovery interval [value ]
no link-flap auto-recovery interval
```

【View】

Global configuration mode, port view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

Command is used to configure the auto-recovery interval of global or port link flapping.

By default, the auto-recovery interval of link flapping is 3600 seconds.

Value: the auto-recovery interval of link flapping, and the value range is < 30-86400 >.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge2
Switch(config-ge2)#link-flap auto-recovery interval 3600
```

3.11.4 Configure Detection Interval of Link Flapping

【Command】

```
link-flap interval [value ]
no link-flap interval
```

【View】

Global configuration mode, port view

【Default Level】

2: Configuration level

【Parameter】

Value: the detection interval of link flapping, and the value range is < 10-100 >.

【Description】

The command is used to configure the detection interval of interface link flapping.

By default, the detection interval of interface link flapping is 20 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge2
Switch(config-ge2)#link-flap interval 20
```

3.11.5 Configure Time Threshold Value of Link Flapping

【Command】

```
link-flap threshold [value ]
no link-flap threshold
```

【View】

Global configuration mode, port view

【Default Level】

2: Configuration level

【Parameter】

Value: link flapping times, and the value range is <3-100>.

【Description】

The command is used to configure the times of interface link flapping.

By default, the time of interface link flapping is 5 times.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge2
Switch(config-ge2)# link-flap threshold 6
```

3.11.6 Check Link Flapping Protection Configuration

【Command】

show link-flap [interface interface-name]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

The command is used to view the configuration and status information of port link flapping protection.

interface-name: specifies the interface type and interface number. Check the information of all ports when no port is specified, and check the information of designated port when port is specified.

【Instance】

```
Switch> enable
Switch# show link-flap interface ge2
```

4 VLAN Configuration

4.1 VLAN Overview

VLAN (Virtual Local Area Network) is a communication technology that logically divides a physical LAN into multiple broadcast domains. Hosts in VLAN can directly communicate with each other, but two VLAN can't directly communicate with each other, which can limit the broadcast message in a VLAN.

Ethernet is a data network communication technology that shares media based on Carrier Sense Multiple Access/Collision Detection (CSMA/CD). When an Ethernet network has a large number of hosts, it can lead to serious collision, broadcast storms and degraded network performance or even result a complete network breakdown. Using switches to connect LANs can mitigate collisions, but cannot isolate broadcast packets or improve network quality.

In this case, VLAN technology appears, which can divide a LAN into multiple logical VLANs, each VLAN is a broadcast domain, and the communication between hosts in VLAN is the same as that in a LAN, but the VLANs cannot communicate with each other directly, so that broadcast messages are limited in one VLAN.

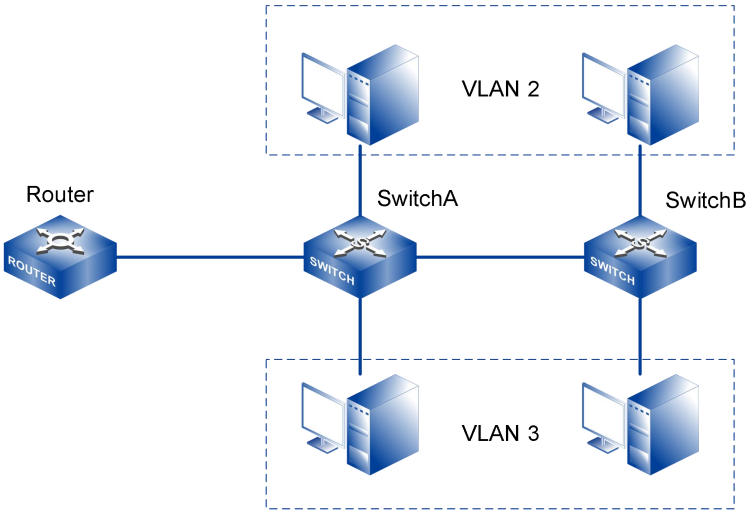


Figure above shows a typical VLAN networking diagram. Two switches are deployed in different locations (for example, on different floors of a building). Each switch is connected to two PCs belonging to different VLANs, which likely belong to different entities or companies.

Using VLAN can bring following benefits to users.

- Limits broadcast domains: broadcast domain is limited in one VLAN. This conserves bandwidth and improves network efficiency.
- Enhances LAN security: packets from different VLANs are transmitted separately. Users in a VLAN cannot communicate directly with users in another VLAN.
- Improves network robustness: fault is limited in one VLAN, it does not affect the normal operation of other VLANs.
- Allows for flexible virtual working group construction: With VLAN technology, different users can be divided into different groups, and users in the same group would not be limited in a fixed physical range, simplifying network construction and maintenance.

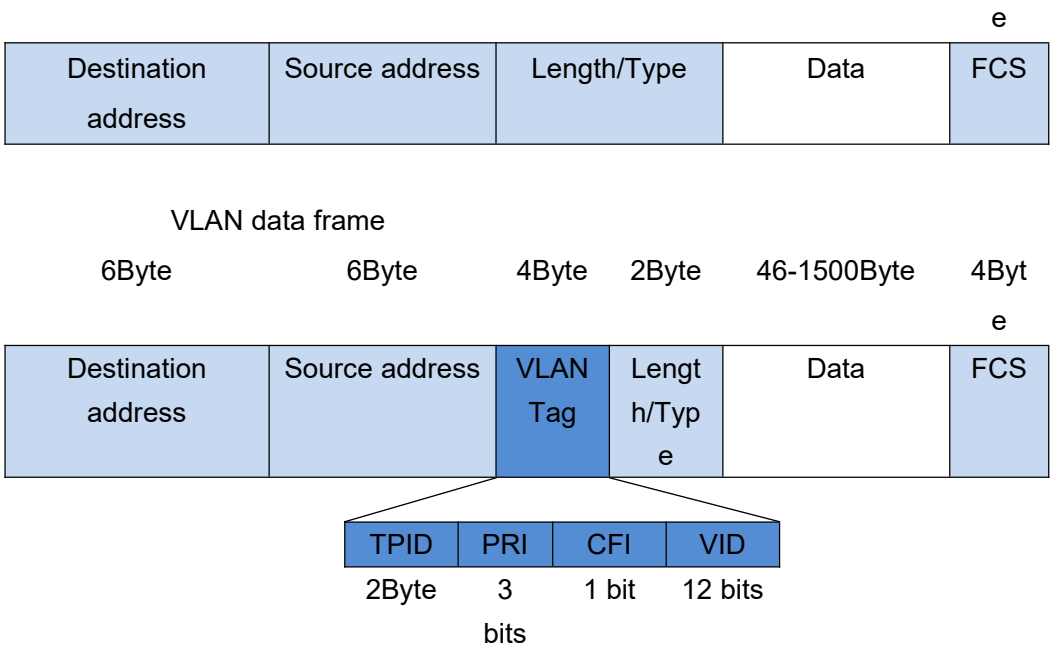
4.2 Principle Description

4.2.1 VLAN Tags

To make switch identify packets from different VLANs, it needs to add fields that identifies the VLAN tags in the message. Under the provisions of IEEE 802.1Q protocol, the device can add 4 bytes VLAN tag (Tag for short) between Source address and Length/Type fields of Ethernet data frame, identifying the VLAN information. As shown below:

Traditional Ethernet data frame

6Byte	6Byte	2Byte	46-1500Byte	4Byt



A VLAN tag contains four fields. The meanings of each field are shown in the following table:

Field	Length	Definition	Value
TPID	2Byte	Tag Protocol Identifier (TPID) indicates the frame type.	Represent the data frame type, when the value is 0x8100, it represents the VLAN data frame of VLAN 802.1Q. An 802.1Q-incompatible device discards this frames. The manufacturers can customize the value of this field. When the neighbor device configures the TPID value to non-0x8100, in order to identify such a message and realize intercommunication, it is necessary to modify the TPID value on this device to ensure that the TPID value configuration is consistent with that of the neighbor device.
PRI	3 bits	Priority (PRI) indicates the	The value ranges from 0 to 7.A

Field	Length	Definition	Value
		802.1p priority of data frame.	larger value indicates a higher priority. During network congestion, the switch will preferentially send data frame with higher priority.
CFI	1 bit	Canonical Format Indicator (CFI) indicates whether a MAC address is encapsulated in canonical format over different transmission media. CFI is used to ensure compatibility between Ethernet and token ring networks.	A CFI value of 0 means that the MAC address is encapsulated in a standard format, and a value of 1 means that it is encapsulated in a non-standard format. In Ethernet, the value of CFI is 0.
VID	12 bits	VLAN ID (VID) indicates the VLAN to which a frame belongs.	VLAN IDs range from 0 to 4095. The values 0 and 4095 are reserved, and therefore valid VLAN IDs range from 1 to 4094.

The switch uses the VID in the VLAN tag to identify the VLAN to which the data frame belongs, and the broadcast frame is only forwarded in the same VLAN, which limits the broadcast domain to one VLAN.

VLAN Tags in Received and Sent Frames：

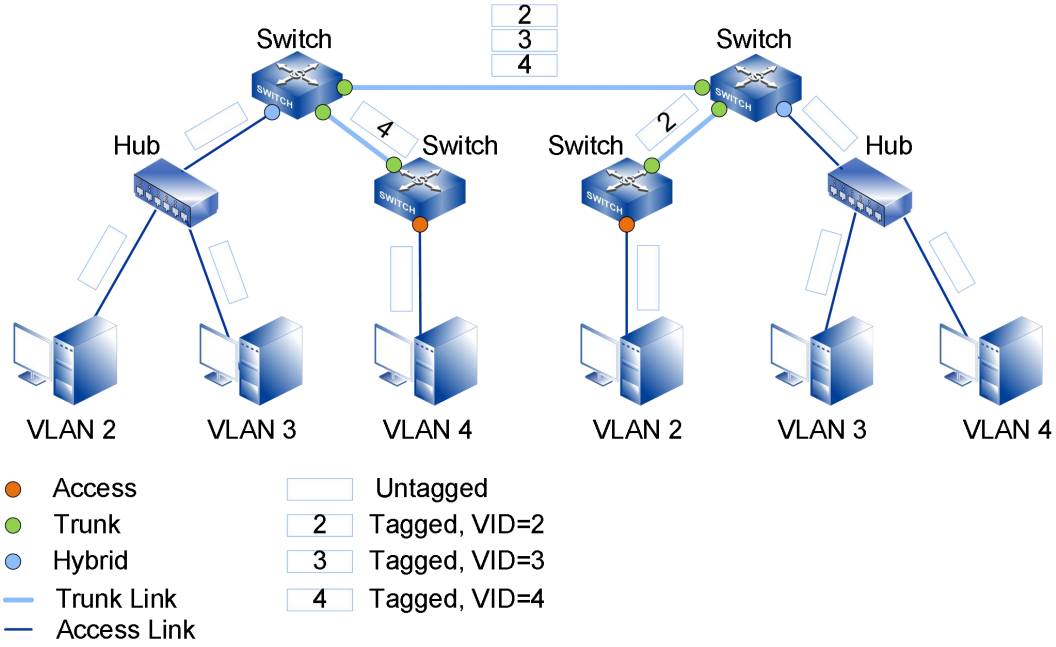
- In a VLAN, Ethernet frames have the following two types:
  - Tagged frame: frame with a 4-byte VLAN tag
  - Untagged frame: frame without a 4-byte VLAN tag
- Common devices:
  - User hosts, servers, hubs, and unmanaged switches can only receive and send untagged frames.
  - Switches, routers, and ACs can receive and send both tagged and untagged frames.
  - Voice terminals and APs can receive and send tagged and untagged frames simultaneously.

All frames processed in a switch carry VLAN tags to improve frame processing efficiency.

4.2.2 Link and Interface Types

All the data frames processed inside the switch are tagged with VLAN, but some devices connected to the switch in the current network can only send and receive Untagged frames. To interact with these devices, it is necessary for the interface to recognize Untagged frames and add and strip VLAN tags to frames when sending and receiving. At the same time, users belonging to the same VLAN in the existing network may be connected to different switches, and there may be more than one VLAN crossing the switches. If intercommunication between users is needed, the interfaces between switches need to be able to identify and send data frames of multiple VLANs at the same time.

In order to adapt to different connections and networking, three interface types, Access Interface, Trunk Interface and Hybrid Interface, and two link types, Access Link and Trunk Link, are defined, as shown in the following figure.



As shown in the above figure, Ethernet links fall into the following types, depending on the number of VLANs to be loaded in links:

- Access link

  An access link can transmit data frames of only one VLAN. It connects a switch to a user terminal, such as a host or server. Generally, user terminals do not need to know the VLANs to which they belong and cannot identify tagged frames; therefore, only untagged frames are transmitted along an access link.

- Trunk link

  A trunk link can transmit data frames from multiple VLANs. It connects a switch to

another switch or a router. Frames on a trunk link must be tagged so that other network devices can correctly identify VLAN information in the frames.

Ethernet interfaces are classified into the following types depending on the objects connected to them and the way they process frames:

- Access interface
  An access interface often connects to a user terminal such as a user host or server that cannot identify VLAN tags, or is used when VLANs do not need to be differentiated. It can receive frames or Untagged frames with the same VLAN ID as PVID, but can only send Untagged frames, and can only add tags of unique VLAN to Untagged frames.
- Trunk interface
  A trunk interface often connects to a switch, router, AP, or voice terminal that can receive and send tagged and untagged frames simultaneously. It allows tagged frames from multiple VLANs and untagged frames from only one VLAN.
- Hybrid interface
  A hybrid interface can connect to a user terminal (such as a user host or server) or network device (such as a hub or simplified Layer 2 switch) that cannot identify tags, but also a switch, router, voice terminal, or AP that can receive and send tagged and untagged frames. It can allow frames with tags from multiple VLANs to pass through, and allow frames sent from such interfaces to be configured with tags from some VLANs (i.e., without stripping tags) and without tags from some VLANs (i.e., stripping tags).

4.2.3 Default VLAN

The default VLAN ID of an interface is called the port default VLAN ID (PVID). Frames processed in a switch all carry VLAN tags.When the switch receives an untagged frame, it adds a VLAN tag to the frame according to the default VLAN of the interface that received the frame. Its specific functions are:

- When an interface receives an untagged frame, the interface adds a tag with the PVID to the frame and sends the frame to the switch for processing. When an interface receives a tagged frame, the tag is not added.
- When an interface sends a frame in which the VLAN ID is the same as the PVID, the switch removes the tag from the frame before sending it out from the interface.

Each interface has a default VLAN. By default, the default VLAN ID of all interfaces is VLAN 1, but users can configure it as needed:

- The default VLAN of an access interface is the VLAN allowed by the access

interface. Changing the allowed VLANs will also change the default VLAN.

- Trunk and hybrid interfaces allow multiple VLANs but have only one default VLAN. Changing the allowed VLANs will not change the default VLAN.

4.2.4 Adding and Removing VLAN Tags

Ethernet data frames are tagged or untagged based on the interface type and default VLAN. The processing methods of data frames for various types of interfaces are shown in the following table.

Interface type.	Process for Receiving Untagged Message	Process for Receiving Tagged Message	The process of transmit frame
Access interface	Receive this message and tag it with default VLAN ID.	• Receive the message when the VLAN ID is the same as default VLAN ID. • Discard the message when the VLAN ID is different from the default VLAN ID.	Removes the PVID tag and transmits the frame.
Trunk interface	• Adds a message with the default VLAN ID and then receives it if the default VLAN ID is in the VLAN ID list that allows to pass. • Adds a message with the default VLAN ID and then discards it if the default VLAN ID is not in the VLAN ID list that allows to pass.	• Receive this message when the VLAN ID is in the list of VLAN ID that allow to pass through the interface. • Discard this message when the VLAN ID is not in the list of VLAN ID that allow to pass through the interface.	• When the VLAN ID is the same as the default VLAN ID, and it is the VLAN ID allowed to pass through the interface, it would strip the Tag and send this message. • When the VLAN ID is different from the default VLAN ID, and it's the VLAN

Interface type.	Process for Receiving Untagged Message	Process for Receiving Tagged Message	The process of transmit frame
			ID allowed to pass through the interface, it would remain its original Tag and send the message.
Hybrid interface	• Adds a tag with the default VLAN ID and then receives it if the default VLAN ID is in the VLAN ID list that allows to pass. • Adds a tag with the default VLAN ID and then discards it if the default VLAN ID is not in the VLAN ID list that allows to pass.	• Receive this message when the VLAN ID is in the list of VLAN ID that allow to pass through the interface. • Discard this message when the VLAN ID is not in the list of VLAN ID that allow to pass through the interface.	When the VLAN ID is the one allowed to pass through the interface, it would send this message. The port can be configured whether to transmit frames with tags.

When the port's entrance filtering function is disabled, the port will be allowed to receive the message of VLAN which is not the port's own, and will forward the message to the designated VLAN.

Interfaces process received frames as follows:

- Access, trunk, and hybrid interfaces add VLAN tags to received untagged frames. Trunk and hybrid interfaces determine whether to accept untagged frames depending on whether VLANs specified by the VLAN IDs in the frames are allowed, whereas an access interface accepts the untagged frames unconditionally.
- Access, trunk, and hybrid interfaces determine whether to accept tagged frames depending on whether VLANs specified by the VLAN IDs in the frames are allowed (the VLAN ID allowed by an access interface is the default VLAN ID).
- When sending data frames:

- An access interface directly removes VLAN tags from frames.
- A trunk interface removes VLAN tags from frames only when their VLAN IDs are the same as the PVID on the interface.
- A hybrid interface determines whether to remove VLAN tags from frames based on the interface configuration.

Frames sent by an access interface are all untagged. On a trunk interface, only frames of one VLAN are sent without tags, and frames of other VLANs are sent with tags. On a hybrid interface, you can specify the VLANs of which frames are sent with or without tags as needed.

4.2.5 VLAN Division

VLAN can be divided based on interface, MAC address, subnet and network layer protocol. Comparison of VLAN division in different ways is shown in the following table.

VLAN Division Modes	Implementation	Advantage, Disadvantage and Usage Scenario
Interface-based	VLANs are assigned based on interfaces. A network administrator preconfigures a PVID for each interface on a switch.When an untagged frame arrives at an interface, the switch adds the PVID of the interface to the frame. The frame is then transmitted in the VLAN specified by the PVID.	Advantage: It is simple to define VLAN members. Disadvantage: The network administrator needs to reconfigure VLANs when VLAN members change. Usage Scenario: Applies to networks of any scale and with devices at fixed locations.
MAC address-based	VLANs are assigned based on source MAC addresses of frames. A network administrator preconfigures mappings between	Advantage: When physical locations of users change, the network administrator

VLAN Division Modes	Implementation	Advantage, Disadvantage and Usage Scenario
	MAC addresses and VLAN IDs.When receiving an untagged frame, the switch adds the VLAN tag mapping the MAC address of the frame to the frame. Then the frame will be transmitted in the specified VLAN.	does not need to reconfigure VLANs for the users. This improves security and access flexibility on a network. Disadvantage: The network administrator must predefine VLANs for all members on a network. Usage Scenario: Applies to small-scale networks where user terminals often change physical locations but their NICs seldom change, for example, mobile computers.
Subnet-based	VLANs are assigned based on the source IP addresses and subnet masks of the frame. A network administrator preconfigures mappings between IP addresses and VLAN IDs.When receiving an untagged frame, the switch adds the VLAN tag mapping the IP address of the frame to the frame. Then the frame is transmitted in the specified VLAN.	Advantage: • When physical locations of users change, the network administrator does not need to reconfigure VLANs for the users. • This mode reduces communication traffic and allows a broadcast domain to span multiple switches. Disadvantage: Users are distributed

VLAN Division Modes	Implementation	Advantage, Disadvantage and Usage Scenario
		regularly and multiple users are on the same network segment. Usage Scenario: Applies to scenarios where there are high requirements for mobility and simplified management and low requirements for security. For example, this mode can be used if a PC with multiple IP addresses needs to access servers on different network segments or a PC needs to join a new VLAN automatically after the PC's IP address changes.
Protocol-based	VLANs are assigned based on protocol (suite) types and encapsulation formats of frames. A network administrator preconfigures mappings between protocol types and VLAN IDs.When receiving an untagged frame, the switch adds the VLAN tag mapping the protocol type of the frame to the frame. Then the frame is transmitted in the specified VLAN.	Advantage: This mode binds service types to VLANs, facilitating management and maintenance. Disadvantage: • The network administrator must preconfigure mappings between all protocol types and VLAN IDs.

VLAN Division Modes	Implementation	Advantage, Disadvantage and Usage Scenario
		• The switch needs to analyze protocol address formats and convert the formats, which consumes excessive resources. Therefore, this mode slows down switch response time. Usage Scenario: Applies to networks using multiple protocols.

If ingress untagged frames match multiple VLAN division modes, such as rule group, the VLAN division modes are selected in descending order of priority: MAC address-based VLAN division> subnet-based VLAN division > protocol-based VLAN division > interface-based VLAN division. If the VLAN is not matched based on MAC, subnet and protocol, the message will be forwarded in the default VLAN of the port.

For Tagged messages, if the VLAN ID carried by the message is in the VLAN list allowed by the port, the message is allowed to pass; otherwise, it is directly discarded.

4.2.6 Intra-VLAN Communication

Packets transmitted between users in a VLAN go through three phases:

- Packet transmission from the source user host
  Before sending a frame, the source host compares its IP address with the destination IP address. If the two IP addresses are on the same network segment, the source host obtains the MAC address of the destination host to fill the destination field MAC address of the frame. If the two IP addresses are on different network segments, the frame needs to be forwarded by the gateway. The source host obtains the gateway's MAC address, and uses it as the destination MAC address.
- Ethernet switching in a switch

The switch determines whether to forward a received frame at Layer 2 or Layer 3 based on the information in the destination MAC address, VLAN ID, and Layer 3 forwarding flag.

- If the destination MAC address and VLAN ID of the frame match a MAC address entry of the switch and the Layer 3 forwarding flag is set. The switch searches for a Layer 3 forwarding entry based on the destination IP address. If no entry is found, the switch searches for a route to forward the frame at Layer 3.
- If the destination MAC address and VLAN ID of the frame match a MAC address entry but the Layer 3 forwarding flag is not set. The switch directly forwards the frame from the outbound interface specified in the matching MAC address entry.
- If the destination MAC address and VLAN ID of the frame do not match any MAC address entry: The switch broadcasts the frame to all interfaces that allow the VLAN specified in the VID in order to obtain the MAC address of the destination host.

- When devices (including switches and user hosts, switches and switches, switches and other network devices) interact with each other, Ethernet switches inside switches are tagged. In order to successfully interact with different devices, switches need to add or strip tags according to interface settings.

After VLANs are assigned, broadcast messages are only forwarded at layer 2 within the same VLAN, users in the same VLAN can directly communicate at Layer 2. There are two intra-VLAN communication scenarios depending on whether hosts in the same VLAN connect to the same or multiple switches.

4.2.7 Inter-VLAN Communication

After VLANs are assigned, broadcast packets are only forwarded in the same VLAN.This means that hosts in different VLANs cannot communicate at Layer 2, thus isolating broadcast. In real-world scenarios, hosts in different VLANs often need to communicate, so inter-VLAN communication needs to be implemented to resolve this. Similar to intra-VLAN communication described in Intra-VLAN Communication, inter-VLAN communication goes through three phases: packet transmission from the source host, Ethernet switching in a switch, and adding and removing VLAN tags during the exchange between devices. According to the Ethernet switching principle, broadcast packets are only forwarded in the same VLAN and hosts in different VLANs

cannot directly communicate at Layer 2. Layer 3 routing or VLAN translation technology is required to implement inter-VLAN communication.

A VLANIF interface is a Layer 3 logical interface that can be used to implement inter-VLAN Layer 3 connectivity. VLANIF is simple to configure, so it is the most commonly used technology for inter-VLAN communication. However, a VLANIF interface needs to be configured for each VLAN and each VLANIF interface requires an IP address. As a result, this technology wastes IP addresses.

In VLANIF interface, users between VLANs can only be in different network segments (because in the same network segment, the host will encapsulate the MAC address of the destination host, and the switch judges to carry out layer 2 switching, which is only in the same VLAN, so broadcast messages cannot reach different VLANs, and the MAC address of the destination host cannot be obtained, so intercommunication cannot be realized).

4.2.8 Intra-VLAN Layer 2 Isolation

You can add different users to different VLANs to implement Layer 2 isolation between users. However, if an enterprise has too many users, then this method needs to assign VLANs for users who can't communicate with each other. This uses a large number of VLANs and increase the configuration and maintenance workload of the network administrator.

Therefore, it provides intra-VLAN Layer 2 isolation technologies including port isolation. Port isolation can isolate interfaces in a VLAN. You can add interfaces to a port isolation group to disable Layer 2 packet transmission between the interfaces.Interfaces in different port isolation groups or are not in any port isolation groups can exchange packets with other interfaces normally. At the same time, the user can configure one-way isolation of the port, creating a more secure and flexible network.

4.2.9 mVLAN

When users centrally manage devices through remote network management, they need to configure IP address on the device through VLANIF interface as device management IP, and connect to the device through management IP using Telnet/SSH protocol for management. If users connected to other interfaces on the device join the VLAN can also log in to the switch. This poses security risks to the switch.

In this situation, VLAN could be configured to management VLAN. In layer 2 switches, users belonging to management VLAN can log in to management device through TELNET, SSH, WEB, SNMP, etc. That is to say, only users belonging to management VLAN can communicate with switch CPU, so layer 2 switches must configure management VLAN. In the layer 3 switch, there is no need to set the management VLAN, because the interface VLAN of the layer 3 switch is the management VLAN.

4.3 Configure VLAN

4.3.1 Enter VLAN Configuration Mode

【Command】

vlan database

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

vlan database: used to enter VLAN configuration mode.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#vlan database
```

4.3.2 Add VLAN ID

【Command】

vlan <vlan-id> （name WORD | ）
vlan range VLANLIST
no vlan <vlan-id>

【View】

VLAN configuration view

【Default Level】

2: Configuration level

【Parameter】

- <vlan-id> : VLAN ID value, range is 2-4094.
- WORD: VLAN name.
- range: set the static VLAN in batch.
- VLANLIST: VLAN range to be set, user can input a single number, continuous range vlan or a combination of single and range, separated by commas, eg: 4, 10-20.

【Description】

vlan: this command is used to create a static VLAN and configure the VLAN name.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#vlan database
Switch(config-vlan)#vlan 2
```

4.3.3 Port Type

【Command】

```
switchport mode (access| hybrid | trunk)
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

- Access: set the link type of the port to access type.
- Hybrid: set the link type of the port to hybrid type.
- trunk: set the link type of the port to trunk type.

【Description】

switchport mode: this command is used to configure the link type of the port.

By default, the link type of all ports are access.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
```

```
Switch(config-ge8)#switchport mode trunk
```

4.3.4 Port Default VLAN

【Command】

```
switchport (access| hybrid ) vlan <vlan-id>
no switchport (access| hybrid ) vlan
switchport trunk native vlan <vlan-id>
no switchport trunk native vlan
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

- access vlan: set the default vlan for a port in access mode to <2-4094>.
- hybrid vlan: sets the default vlan for a port in hybrid mode to <2-4094>.
- <2-4094>: VID allowed setting range is 2-4094.
- Native vlan: set the local VLAN and classify the unmarked traffic through the Layer 2 interface, that is, set the PVID of the port.

【Description】

- **switchport (access | hybrid) vlan:**
  The command is to reset the default VLAN of the port. For example, enter the configuration mode of port ge1, the port ge1 is in hybrid mode, and enter “switchport hybrid vlan 3”. The default VLAN ID for port ge1 becomes 3.
- **switchport trunk native vlan:**
  This command specifies a local VLAN for a Trunk port. As a Trunk port, it must belong to a native VLAN. The native VLAN refers to UNTAG messages sent and received on the interface, which are all considered to belong to the VLAN. Obviously, the default VLAN ID of the interface (i.e., PVID in IEEE 802.1Q) is the VLAN ID of the native VLAN. At the same time, if native VLAN frames are sent in Trunk port, UNTAG must be adopted.

By default, the default VLAN ID of the port is 1.

Notice

When setting the VLAN ID, the port mode parameter of the command must be consistent with the current mode of the setting port to ensure the setting takes effect, otherwise an error

will be returned.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#switchport access vlan 2
```

4.3.5 Classify VLAN Based on Port

【Command】

```
switchport hybrid allowed vlan add (tag| untag )<vlan-id>
switchport hybrid allowed vlan remove <vlan-id>
switchport hybrid allowed vlan (all | none)
switchport trunk allowed vlan (add |except | remove) <vlan-id>
switchport trunk allowed vlan (all | none)
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

- <vlan-id> : VLAN ID, range is 1-4094.
- add: add the port to the VLAN.
- all: add ports to all VLANS.
- except: adds a port to all VLANS except the one specified.
- none: delete the port from all VLANS except PVID.
- remove: delete the port from the specified VLAN.
- tag: the port will add a VLAN tag when forwarding a VLAN message.
- untag: the port will remove the VLAN Tag when forwarding a VLAN message.

【Description】

switchport (hybrid | trunk) allowed vlan: this command is used to configure the port to be added to or removed from a specified VLAN.

Notice

- When adding hybrid or trunk ports to a VLAN, the port should be set to the appropriate type.

- Hybrid or trunk ports are untag in the VLAN to which PVID belongs, and trunk ports are tag in VLAN except PVID.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#switchport hybrid allowed vlan add tag 2
```

4.3.6 Port Receive Frame Type

【Command】

switchport acceptable-frame-type (all | tagged | untagged)

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

- All: there is no restriction on whether the received message with tag.
- Tagged: only allow the port to receive tagged message, that is, stop the port to receive untagged message.
- untagged: only allow the port to receive untagged message, stop the port to receive tagged message.

【Description】

switchport accept-frame-type: this command is used to restrict whether the port is allowed to accept message with tags.

By default, the accepted frame type of the port is set to all, which means that the port does not by default restrict whether or not it can receive packets with tags. However, if the accepted frame type of the port is set to tagged, the port can only allow tagged message to pass, and other packets will be discarded.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#switchport acceptable-frame-type tagged
```

4.3.7 Port Entry Filtering

【Command】

switchport ingress-filter (enable | disable)

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

- enable: only messages from the VLAN to which the port belongs are allowed to be received.
- disable: allows to receive messages from VLANs that do not belong to the port.

【Description】

The ingress filter function of the device defaults to enable, that is, any port only allows packets belonging to its VLAN to pass through. Other message will be discarded. However, when the ingress filter function of the port is set to disable, the port will allow to receive messages not belonging to the VLAN of the port and forward the message to the specified VLAN.

For example, set port entry filtering on port ge1 to disable. Port ge1 belongs to vlan10. The message is sent to ge2 belonging to vlan20. After ge1 receives the message, the port will receive the message and forward the message to the specified vlan. In other words, the message will be forwarded to ge2 belonging to vlan20 instead of discarding the message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#switchport ingress-filter disable
```

4.3.8 VLAN Division Based on Subnet/MAC/ Protocol

4.3.8.1 VLAN Classifier Function Introduction

The VLAN classifier is similar to PVID in that it assigns a default VLAN ID to packets entering the switch port. It provides MAC-based, subnet-based and protocol-based

assignment. If a packet matches all of the three, only one rule will take effect. The priority of the rules is: MAC-based, subnet-based, and protocol-based. For example, if MAC-based rules are matched first, neither subnet-based nor protocol-based VLAN assignment rules will take effect. If none of the three rules match, VLAN ID are assigned according to PVID rule.

4.3.8.2 Classify VLAN Based on Sub-network

【Command】

vlan classifier rule <1-256> ipv4 A.B.C.D/M vlan <1-4094>

【View】

Global config

【Default Level】

2: Configuration level

【Parameter】

- <1-256>: group number.
- A.B.C.D/M: sub-network segment.
- <1-4094>: represents the VLAN ID assigned by the matching rule.

【Description】

Configure subnet-based VLAN rules.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#vlan classifier rule 1 ipv4 192.168.2.0/24 vlan 2
```

4.3.8.3 Classify VLAN Based on MAC Address

【Command】

vlan classifier rule <1-256> mac WORD vlan <1-4094>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

- <1-256>: group number.

- MAC: MAC address.
- <1-4094>: represents the VLAN ID assigned by the matching rule.

【Description】

Configure MAC-based VLAN rules.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#vlan classifier rule 1 mac 0.0.2 vlan 3
```

4.3.8.4 Classify VLAN Based on Protocol

【Command】

```
vlan classifier rule <1-256> proto
(ip|ipv6|ipx|x25|arp|rarp|atalkddp|atalkaarp|atmmulti|atmtrans
port|pppdiscovery|pppsession|xeroxpup|xeroxaddrtrans|g8bpqx25|
ieeepup|ieeeaddrtrans|dec|decdnadumpload|decdnaremoteconsole|d
ecdnarouting|declat|decdiagnostics|deccustom|decsyscomm|<0-655
35>) encap (ethv2|snapllc|nosnapll) vlan <1-4094>
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

- <1-256>: group number.
- proto：Ethernet protocol type; enter ip, ipv6, ipx, x25, arp, rarp, atalkddp, atalkaarp, atmmulti, atmtransport, pppdiscovery, pppsession, xeroxpup, xeroxaddrtrans, g8bpqx25, ieeepup, ieeeaddrtrans, dec, decdnadumpload, decdnaremoteconsole, decdnarouting, declat, decdiagnostics, deccustom, decsyscomm or enter protocol number <0-65535>
- encap：Ethernet Encapsulation Type；ethv2，snapllc，nosnapll.
- <1-4094>: represents the VLAN ID assigned by the matching rule.

【Description】

Configure protocol-based VLAN rules.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#vlan classifier rule 1 proto ip encap ethv2 vlan
3
```

4.3.8.5 Delete VLAN Rule

【Command】

no vlan classifier rule <1-256>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<1-256>: group number.

【Description】

Delete a rule.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#no vlan classifier rule 1
```

4.3.9 Configure VLAN Classification Group

【Command】

vlan classifier group <1-16> (add | delete) rule <1-256>

no vlan classifier group <1-16>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

- <1-16>: group number.
- add: add a rule to a group.
- delete: delete a rule from a group.
- <1-256> : rule number;

【Description】

vlan classifier group: group configuration.

no vlan classifier group: delete group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#vlan classifier group 1 add rule 2
Switch(config)#no vlan classifier group 2
```

4.3.10 Configure the Interface VLAN Classification Group

【Command】

```
vlan classifier activate <1-16>
no vlan classifier activate <1-16>
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

<1-16> : reference group number;

【Description】

vlan classifier activate: interface reference group.

no vlan classifier activate: delete interface reference group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#vlan classifier activate 1
Switch(config-ge1)#no  vlan classifier activate 2
```

4.3.11 Display VLAN Information

【Command】

```
show vlan (<1-4094> | all | auto | brief | dynamic | static |
static-ports)
show vlan classifier group <1-16>
show vlan classifier group interface IFNAME
show vlan classifier interface group <1-16>
show vlan classifier rule <1-256>
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

- <1-4094> : VLAN ID range 1-4094.
- <1-16> : Group ID range is 1-16.
- IFNAME: port name.
- <1-256>：VLAN classification rule ID, range 1-256.

【Description】

show vlan (<1-4094> | all | auto | brief | dynamic | static | static-ports)：

- <1-4094> : displays the specified VLAN information.
- all: displays all VLAN information.
- auto: displays automatically configured VLAN information.
- brief: displays VLAN information of all bridges.
- dynamic: displays dynamic VLAN information.
- static: display static VLAN information.
- Static-ports: display static VLAN egress port/disabled port information.

show vlan classifier group <1-16>： displays the rule ID information of specified VLAN classification group.

show vlan classifier group interface IFNAME： Displays the VLAN classification group information of the specified port is currently running.

show vlan classifier interface group <1-16>： Displays port information for running VLAN classification groups.

show vlan classifier rule <1-256>： displays the VLAN classification information of specified rules.

【Instance】

```
Switch> enable
Switch# show vlan 2
```

5 Ring Configuration

5.1 Ring Overview

Redundant links are usually used in Ethernet switching networks in order to backup links and improve network reliability. However, the use of redundant links will lead to loops on the switching network, causing broadcast storms and unstable MAC address tables and resulting in poor communication quality and even interruption of communication. In order to solve the loop problem in switching network, spanning tree protocol is proposed. However, due to the slow convergence speed of spanning tree protocol topology, it can not meet the needs of industrial control network, so our company provides a self-developed Ring private loop network protocol solution.

Note

Private protocols can't communicate with other manufacturers' devices.

Ring is an Ethernet Ring network algorithm developed and designed for highly reliable industrial control network applications that require link redundancy backup. its design concept is completely in accordance with international standards (RSTP and RSTP) implementation, and do the necessary for industrial control application optimization, with Ethernet link redundancy, fault fast automatic recovery ability.

Ring adopts the design of no master station. The devices running the Ring protocol discover the loop in the network by exchanging information with each other, and block a certain port. Finally, the ring network structure is trimmed into a tree network structure without loop, thus preventing messages from circulating continuously in the ring network, and avoiding the reduction of processing capacity caused by repeated reception of the same message. In a multi-Ring network composed of 250 switches,

when the network is interrupted or fails, the ring can ensure that the user network automatically resumes link communication within 20 ms.
Ring needs to manually divide the ring network ports in advance, support multiple ring network types such as single ring, coupled ring, chain and Dual Homing, and provide visual management of network topology. In a single Ring, Ring supports master/slave and no master configuration to meet various network environment requirements.

5.2 Principle Description

After the Ring Ring network is divided, the Ring protocol configures the devices on the ring network as nodes with different roles on the ring ring, and detects the ring network status and transmits the ring network topology change information through Hello messages between each

network identity node. The nodes on the ring block or release their ports in time according to the state of the ring network, which can eliminate the loop when the ring network is formed; When the equipment or link on the ring network fails, the backup link can be enabled quickly to ensure the smooth operation of the business.

5.2.1 Network ID

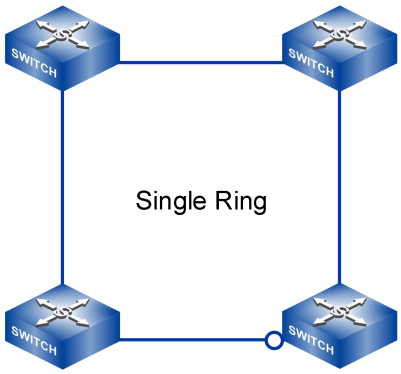
When multiple switches form a ring, the current ring ID would be network ID. Different ring network has different ID. The ring network identification must remain the same in one ring network.

5.2.2 Ring port

Port that can be used for the formation of ring network in switch. In a coupling ring, the coupling port is the port connected to the different network identifiers, and the control port is the port in the link where the two rings meet.

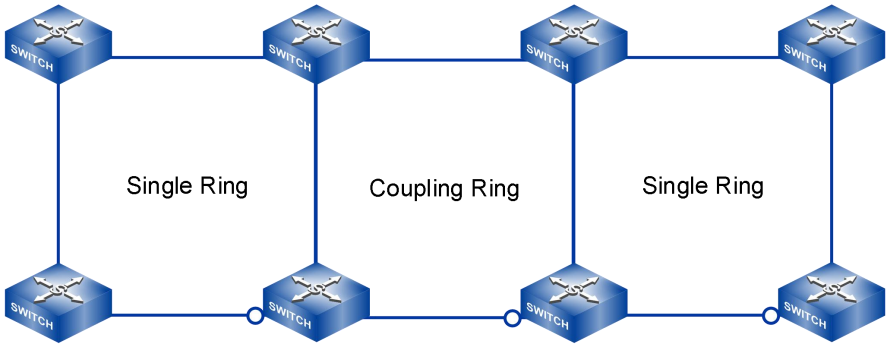
5.2.3 Ring Type

- Single Ring



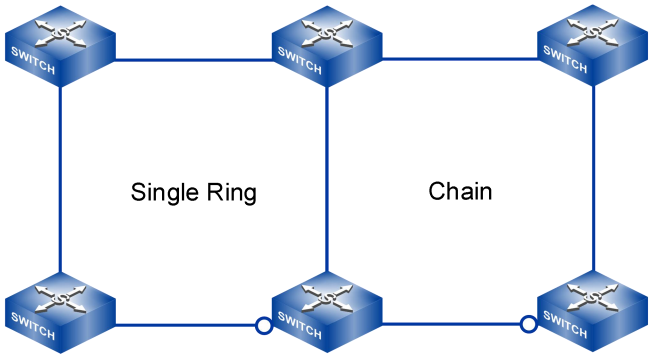
A Single Ring is a basic ring networking structure in which all devices are connected in a ring. When the network is working normally, the algorithm running on the device will automatically block a link as a backup link to ensure the normal operation of the network. When the network has a link failure, the algorithm will automatically start the backup link and resume data communication within 20 ms.

- Coupling Ring



Coupling Ring is a redundant structure introduced to connect two separate networks. The Coupling Ring provides additional security by enabling the coupling of two ports on different switches. For some systems, users can also create single rings for devices from different regions and also integrate multiple single rings through the Coupling Ring to create a larger redundant network.

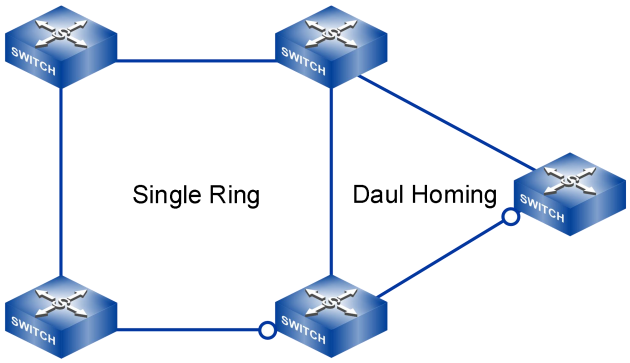
- Chain



Chain refers to connecting multiple switch devices in series and connecting both

ends of the Chain to Ethernet network. Chain has strong channel selection ability. When the network is working properly, the algorithm automatically blocks one link in the Chain as a backup link, forcing all devices to access the Ethernet network from the unblocked end of the link. When Chain link failure occurs, the algorithm will automatically start the backup link within 20ms and quickly guide the device access the Ethernet network through the side that do not has a link failure.

- Dual Homing



Dual Homing is a special case of the Chain, in which users can host the same switch on two different networks or two different switching devices on the same network. The algorithm will automatically select one link for data communication according to the link condition. When the link in the communication state fails, the other link will start to work within 20ms.

Notice:

- RING loop ports can be normal physical ports or static aggregation groups.
- The RING loop port cannot enable other layer 2 protocols (STP/RSTP/MSTP, ERPS, etc.) at the same time.

5.2.4 Master/Slave Mode

In the single ring network, the ring network supports two network architectures: no master station and master station.

- No master station: Only when all devices are configured in Slave mode, the single ring has no master station structure, that is, there is no designated backup link. In the ring network without master station, when the ring network fails, the network enables the backup link to ensure the normal operation of the network; When the failed link is recovered, the network will not switch the backup link, and the recovered link will be taken as the current backup link.

- Master station: when a device is designated to be configured in master mode, the device serves as the master device, and the single-ring network in which it is located has a master station structure; Other single-ring devices need to be configured as Slave devices. In the single-ring network, the link connected with "ring network port 2" in the main device "single ring" is used as the designated backup link. In the ring network with master station, when the ring network fails, the network enables the backup link to ensure the normal operation of the network; When the failed link is restored, the network switches back to the link connected with the designated master device as the backup link.

Note

- Master-slave mode is only applicable to single Ring network at present, and other Ring networks are no-master station structure.
- Master-slave mode is recommended to be configured as a master-slave mode, in which one device is designated as the master device and other devices are designated as the slave devices.
- When Master-mode devices appears in a single ring, the ring network is master station structure. When multiple Master mode devices appear in the configuration, the system will select one Master mode device as the master device according to the specified algorithm.

5.3 Ring Configuration

5.3.1 Global Ring Enablement

【Command】

`[no] ring`

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

`ring`: this command is used to enable the global Ring function.

`no ring`: this command is used to disable the global Ring function and delete all Ring groups.

By default, the global Ring function is disabled.

【Instance】

```
Switch> enable
Switch# configure terminal
Switch(config)# ring
```

5.3.2 Create Ring NetworkGroup

【Command】

```
ring <group-id> id <ring-id> port1 <ifname> port2 <ifname> type
0 hello <hello-time> (master | slave)
ring <group-id> id <ring-id> port1 <ifname> port2 <ifname> type
<type-id> hello <hello-time>
no ring <group-id>
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

- group-id: ring group ID, range is 1-12.
- ring-id: ring loop ID, range is 0-255.
- ifname: ring port name, the port can be a physical port or a static aggregation group, and the port cannot enable spanning tree or ERPS.
- type-id: ring loop type, range 0-3, corresponding to Single Ring, Coupling Ring, Chain, Dual Homing.
- hello-time: hello request packet sending period, range 0-300(*100ms), 0 means to do not send.
- Master | slave: ring network master device selection, no master station if all are masters, only ring type is Single ring can be configured.

【Description】

ring <group-id>: this command is used to configure the ring group.

no ring <group-id>: this command is used to delete the ring group.

By default, no ring group is enabled.

【Instance】

```
Switch> enable
Switch# configure terminal
Switch(config)# ring 1 id 1 port1 ge1 port2 ge3 type 1 hello 1
```

5.3.3 Display Ring Network Information

【Command】

show ring [<*group-id*>]

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

group-id : ring group ID, range is 1-12

【Description】

show ring: this command is used to show ring information.

【Instance】

```
Switch(config)#interface ge3
Switch(config-ge3)#spanning-tree disable
*Switch(config-ge3)# exit
*Switch(config)#interface ge4
*Switch(config-ge4)#spanning-tree disable
*Switch(config-ge4)#exit
*Switch(config)#ring 1 id 1 port1 ge3 port2 ge4 type 0 hello 0 master
*Switch#show ring
ring global : Enable
ring list:
ring GROUP: 1
ring ID: 1
ring PORT1: ge3
ring PORT1 state: block
ring PORT2: ge4
ring PORT2 state: block
ring TYPE: Single
ring HELLOTIME: 0
ring : master
```

6 STP/RSTP/MSTP Configuration

6.1 STP/RSTP/MSTP Overview

6.1.1 STP/RSTP Overview

Redundant links are usually used in Ethernet switching networks in order to backup links and improve network reliability. However, the use of redundant links will lead to loops on the switching network, causing broadcast storms and unstable MAC address tables and resulting in poor communication quality and even interruption of communication. To solve the loop problem in switching network, Spanning Tree Protocol (STP) is proposed.

Devices running STP exchange information to discover loops on the network and block some ports. The ring network structure is pruned into tree network structure without loop to prevent messages from looping in ring network and that the packet processing capabilities of switches is not impacted by receiving the same messages. Due to the slow convergence rate of STP topology, IEEE published the 802.1w standard in 2001 to define RSTP (Rapid Spanning Tree Protocol). RSTP has made improvement on the basis of STP, which has achieved quick topological convergence of network.

After a spanning tree protocol is configured on an Ethernet switching network, the protocol calculates the network topology to implement the following functions:

- Loop prevention: The spanning tree protocol blocks redundant links to prevent potential loops on the network.
- Link redundancy: If an active link fails and a redundant link exists, it will activate the redundant link to ensure network connectivity.

6.1.2 DHCP Overview

RSTP has made improvement on the basis of STP, which has achieved quick topological convergence of network. However, RSTP and STP still have the same defect: because all VLANs in the LAN share a spanning tree, it is impossible to balance the load of data traffic among VLANs, and if the link is blocked, it will not carry any traffic, and some VLAN messages may not be forwarded. To remedy the defects of STP and RSTP, IEEE 802.1s standard issued by IEEE in 2002 has defined Multiple Spanning Tree Protocol (MSTP). To address the limitation of STP and RSTP, MSTP allows fast convergence and provides multiple paths to load balance VLAN traffic.

The following functions could be realized by deploying MSTP protocol in Ethernet:

- Form multiple trees without loops to solve the broadcast storm and realize redundant backup.
- Multiple spanning trees realize load balancing among VLANs, and traffic of different VLANs is forwarded according to different paths.

6.2 Principle Description

6.2.1 STP Principle Description

6.2.1.1 Root Bridge

A tree-shaped network structure must have roots, so STP introduced the concept of Root Bridge. There is only one root bridge on the entire STP network. The root bridge is the logical center, but not necessarily the physical center, of the network. The root bridge changes dynamically with the network topology. After the network converges, the root bridge will generate and send out the configuration BPDU according to a certain time interval, and other devices will only process the message and transmit the topology change record, thus ensuring the stability of the topology.

6.2.1.2 Two Metrics

A spanning tree is calculated based on the following metrics: ID and path cost.

IDs are classified into bridge ID (BID) and port ID (PID).

- BID：
  - According to the IEEE 802.1D standard, a BID is composed of a bridge priority (leftmost 16 bits) and a bridge MAC address. The BID bridge priority

occupies the upper 16 bits, and the remaining lower 48 bits are MAC addresses.

- On an STP network, the device with the smallest BID acts as the root bridge.

- PID：
  - A PID is composed of a port priority (leftmost 4 bits) and a port number (rightmost 12 bits).
  - The PID is used to select the designated port.

Notes:

The port priority affects the role of a port in a spanning tree instance.

The path cost is a port variable used by STP protocol for link selection as a reference value. STP calculates path costs to select effective links, block redundant links, and trim the network into a loop-free tree topology.

On an STP network, a port's path cost to the root bridge is the sum of the path costs of all ports between the port and the root bridge. This path cost is called the root path cost.

6.2.1.3 Three-element Election

Three elements are involved in trimming a ring network into a tree network: root bridge, root port, and designated port.

- Root Bridge (RB): the root bridge is the bridge with the smallest BID as determined by exchanging configuration BPDU.
- Root Port (RP): the root port on an STP device is the port with the smallest path cost to the root bridge and is responsible for forwarding data to the root bridge. The selection criteria for this port are based on the root path cost. An STP device has only one root port, and there is no root port on the root bridge.
- Designated Port (DP): In a LAN, the designated bridge port responsible for forwarding configuration messages to the network segment.

After the root bridge, root ports, and designated ports are selected successfully, a tree topology is set up on the entire network. When the topology is stable, only the root port and designated ports forward traffic. The other ports are in Blocking state; they only receive STP BPDUs and do not forward user traffic.

6.2.1.4 Four Comparison Principles

During role election, STP devices compare the four fields of a BPDU priority vector root ID, root path cost, sender BID, PID. The main information of the port to carry in a configuration BPDU are as follows:

- Root ID： ID of the root bridge.
- Root Path Cost: Path cost to the root bridge is determined by the distance between the port sending the configuration BPDU and the root bridge.
- Sender BID： BID of the device that sends the configuration BPDU.
- Sending port PID: PID of the port that sends the configuration BPDU.

After a device on the STP network receives a configuration BPDU, the four comparison principles are as follows:

- Smallest BID: used to select the root bridge. Devices on an STP network select the device with the smallest BID based on the root ID field.
- Smallest root path cost: used to select the root port on a non-root bridge. On the root bridge, the path cost of each port is 0.
- Smallest sender BID: used to select the root port from ports with the same root path cost. The port with the smallest BID is selected as the root port in STP calculation.
- Smallest PID: used to determine which port should be blocked when multiple ports have the same root path cost.

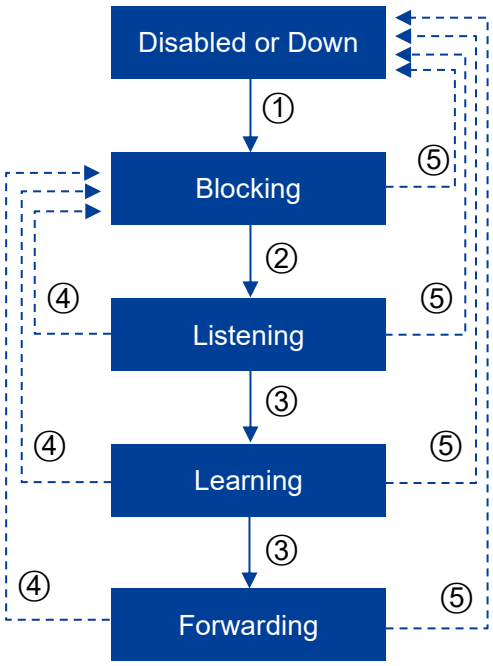
6.2.1.5 Five Port States

The states of ports on an STP device are as follows.

Port state	Destination	Note
Forwarding	Port can forward user traffic and process BPDUs.	Only the root port and designated port can enter the Forwarding state.
Learning	The device creates MAC address entries based on user traffic received on the port but does not forward user traffic through the port.	This is a transitional state, which is designed to prevent temporary loops.
Listening	All ports are in Listening state before the root bridge, root port, and designated port are selected.	This is a transitional state.

Port state	Destination	Note
Blocking	A port receives and processes only BPDUs, and does not forward user traffic.	This is the final state of a blocked port.
Disabled	A port in Disabled state does not process BPDUs or forward user traffic.	The port is Down.

STP port state migration:



①: The port is initialized or enabled, and enters the blocking state

②: The port is selected as the root port or designated port, and enters the listening state

③: The temporary state stay time of the port is up and enters the next state.

The port (learning state or forwarding state) is selected as the root port or designated port

④: The port is no longer the root port, the designated port or the designated state, and enters the blocking state

⑤: port is disabled or link is down

The following parameters affect the STP port states and convergence.

- Hello Time

  The time interval between devices running STP protocol to send configuration message BPDU is used for the devices to detect whether the link is faulty. the device sends BPDU messages to surrounding devices at every Hello Time to confirm whether the link is faulty.

  When the Hello Time is changed, the new value takes effect only after a new root bridge is elected. The new root bridge includes the new Hello Time value in BPDUs it sends to non-root bridges. If the network topology changes, TCN BPDUs are immediately transmitted regardless of the Hello Time.

- Forward Delay

  The Forward Delay timer specifies the length of delay before a port state transition. Link failure will cause the network to recalculate spanning tree, and the structure of spanning tree will change accordingly. However, new configuration BPDUs cannot be immediately spread over the entire network. If the new root port and designated port forward data immediately, transient loops may occur. Therefore, STP defines a port state transition delay mechanism. The newly selected root port and designated port must wait for two Forward Delay intervals before transitioning to the Forwarding state. During this period, the new configuration BPDUs can be transmitted over the network, preventing transient loops.

  The default Forward Delay timer value is 15 seconds. This means that the port stays in Listening state for 15 seconds and then stays in Learning state for another 15 seconds before transitioning to the Forwarding state. The port is blocked when it is in Listening or Learning state, effectively preventing transient loops.

- Max Age

  The Max Age specifies the aging time of BPDUs. This parameter is configurable on the root bridge.

  The Max Age is spread to the entire network with configuration BPDUs. After a non-root bridge receives a configuration BPDU, it compares the Message Age value with the Max Age value in the received configuration BPDU.

  - If the Message Age value is smaller than or equal to the Max Age value, the non-root bridge forwards the configuration BPDU.
  - If the Message Age value is greater than the Max Age value, the non-root bridge discards the configuration BPDU. When this happens, the network

size is considered too large and the non-root bridge disconnects from the root bridge.

If the configuration BPDU is sent from the root bridge, the Message Age value is 0. Otherwise, the Message Age value is the total time spent to transmit the BPDU from the root bridge to the local bridge, including the transmission delay. The Message Age value of a configuration BPDU increases by 1 each time the configuration BPDU passes through a bridge.

6.2.1.6 STP Message Format

Bridge ID, path overhead, port ID and other information, all of which are transmitted through BPDU protocol messages.

Configuration BPDUs are heartbeat packets. STP-enabled designated ports send configuration BPDUs at Hello timer intervals.

Topology Change Notification (TCN) BPDUs are sent only after a device detects a network topology change.

BPDU message is encapsulated in Ethernet frame. Its destination MAC address is a multicast MAC address, 01-80-C2-00-00-00. The Length/Type field is the MAC data length, then followed by LLC header, followed by BPDU message header. The format of Ethernet data frame is as follows.

6 bytes	6 bytes	2 bytes	3 bytes	38-1492 bytes	4 bytes
DMAC	SMAC	Length	LLC	BPDU Data	CRC

Configuration BPDUs are the most common type of BPDU and are sent to exchange topology information among STP devices.

Each bridge actively sends configuration BPDUs during initialization. However, after the network topology is stable, only the root bridge actively sends the configuration BPDU, and other bridges trigger sending their own configuration BPDU after receiving the configuration BPDU from upstream. A configuration BPDU is at least 35 bytes long and includes parameters such as the BID, root path cost, and PID. A bridge processes a received configuration BPDU only if either the sender BID or PID is different from that on the local bridge receive port. If both fields are the same as those on the receive port, the bridge discards the configuration BPDU. Therefore, the bridge does not need to process BPDUs with the same information as the local port.

A configuration BPDU is sent in one of the following scenarios:

- After STP is enabled on ports of a device, the designated port on the device sends configuration BPDUs at Hello timer intervals.

- When the root port receives the configuration BPDU, the device where the root port is located will copy the configuration BPDU to each designated port.
- When a designated port receives an inferior configuration BPDU, the designated port immediately sends its own configuration BPDU to the downstream device.

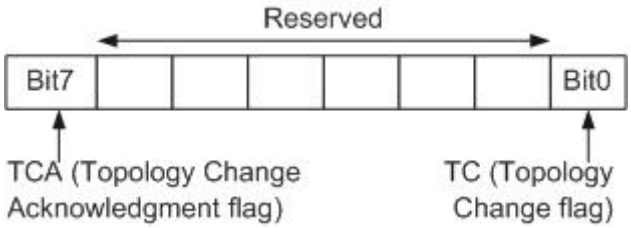
The basic format of BPDU message:

Domain	Bytes	Note
Protocol Identifier	2	The value is fixed at 0, representing a spanning tree protocol.
Protocol Version Identifier	1	The value is fixed at 0, representing a spanning tree protocol.
BPDU Type	1	Indicates the type of a BPDU. The value is one of the following: • 0x00: configuration BPDU • 0x80: TCN BPDU
Flags	1	Indicates whether the network topology has changed. • The rightmost bit is the Topology Change (TC) flag. • The leftmost bit is the Topology Change Acknowledgment (TCA) flag.
Root Identifier	8	Indicates the BID of the current root bridge.
Root Path Cost	4	Indicates the accumulated path cost from a port to the root bridge.
Bridge Identifier	8	Indicates the BID of the bridge that sends the BPDU.
Port Identifier	2	Indicates the ID of the port that sends the BPDU.
Message Age	2	Records the time that has elapsed since the original BPDU was generated on the root bridge. If the configuration BPDU is sent from the root bridge, the Message Age value is 0. Otherwise, the Message Age value is the total time spent to transmit the BPDU from the root

Domain	Bytes	Note
		bridge to the local bridge, including the transmission delay. The Message Age value of a configuration BPDU increases by 1 each time the configuration BPDU passes through a bridge.
Max Age	2	Indicates the aging time of a BPDU.
Hello Time	2	Indicates the interval at which BPDUs are sent.
Forward Delay	2	Indicates the period during which a port stays in Listening and Learning states.

The flag field is shown in the following figure, and only the highest and lowest bits are used in STP.

Format of the Flags field



TCN BPDU

A TCN BPDU contains only three fields: Protocol Identifier, Version, and Type, as shown in Table 1. The Type field is four bytes long and is fixed at 0x80.

When the network topology changes, TCN BPDUs are transmitted upstream until they reach the root bridge. A TCN BPDU is sent in either of the following scenarios:

- A port transitions to the Forwarding state.
- A designated port receives a TCN BPDU and sends a copy to the root bridge.

6.2.1.7 STP Topology Calculation

After STP is enabled on all devices on a network, all devices initially consider themselves as the root bridge. They only transmit and receive BPDUs and do not forward user traffic, and all ports on the devices are in Listening state. The devices

select the root bridge, root ports, and designated ports based on configuration BPDUs.

STP Algorithm Implementation:

- Initialization

  Because each bridge considers itself the root bridge, the BPDU sent from a port is set as follows: The root ID is the BID of the local bridge, the root path cost is the accumulative path cost from the port to the local bridge, the sender BID is the BID of the local bridge, and the PID is the ID of the port that sends the BPDU.

- Root bridge election

  When the network is initialized, all STP devices in the network think of themselves as "Root Bridges", and the root bridge ID is their device ID. Then devices exchange configuration BPDUs and compare their root IDs to find the device with the smallest BID, which becomes the root bridge.

- Root port and designated port selection

  1） The non-root bridge device identifies the port that receives the optimal configuration message as the root port

  - Each port compares the received configuration message with its own configuration message: if the priority of the received configuration message is low, it will be directly discarded and its own configuration message will not be processed; If the priority of the received configuration message is higher, replace the content of the self-configuration message with the content of the configuration message.
  - The device compares configuration messages for all ports and selects the best configuration message

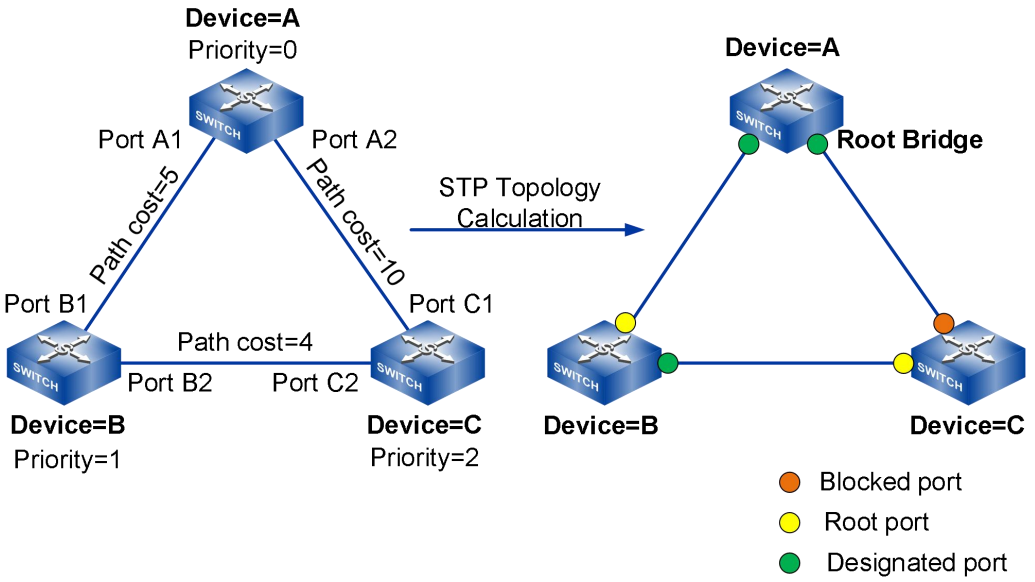
  2） According to the configuration message of the root port and the path cost of the root port, the device calculates a designated port configuration message for each port:

  - Replaces the root ID with the root ID in the configuration BPDU on the root port.
  - Replaces the root path cost with the sum of the root path cost in the configuration BPDU on the root port and the path cost of the root port.
  - Replaces the sender BID with the local BID.
  - Replaces the PID with the local port ID.

  3） The device compares the calculated configuration BPDU with the configuration BPDU received on the port:

- If the calculated configuration message is better, the port is determined as the designated port, and its configuration message is also replaced by the calculated configuration message and sent out periodically.
- If the port's own configuration message is better, it won't update the port's configuration message and block the port. This port will no longer forward data and will only receive and not send configuration messages.

After the root bridge, root ports, and designated ports are selected successfully, a tree topology is set up on the entire network. The following example illustrates how STP calculation is implemented.



DeviceA, DeviceB, and DeviceC are deployed on the network, with priorities 0, 1, and 2, respectively. The path costs between DeviceA and DeviceB, DeviceA and DeviceC, and DeviceB and DeviceC are 5, 10, and 4, respectively.

Table below lists the initial state of each device.

Device	Port Name	Configuration BPDU
DeviceA	Port A1	{0, 0, 0, Port A1}
	Port A2	{0, 0, 0, Port A2}
DeviceB	Port B1	{1, 0, 1, Port B1}
	Port B2	{1, 0, 1, Port B2}
DeviceC	Port C1	{2, 0, 2, Port C1}
	Port C2	{2, 0, 2, Port C2}

The specific meaning of each item in the configuration message is: {root bridge ID, cumulative root path overhead, sender BID, sending port PID}.

Table below describes comparison process and result of each device.

Device	comparison process	Configuration BPDU After Comparison
DeviceA	• Port A1 receives the configuration BPDU {1, 0, 1, Port B1} from Port B1 and finds it inferior to its own configuration BPDU {0, 0, 0, Port A1}, so Port A1 discards the received configuration BPDU. • Port A2 receives the configuration BPDU {2, 0, 2, Port C1} from Port C1 and finds it inferior to its own configuration BPDU {0, 0, 0, Port A2} superior, so Port A2 discards the received configuration BPDU. • DeviceA finds that the root bridge and designated bridge specified in the configuration BPDUs on its ports are on itself. Therefore, DeviceA considers itself as the root bridge and periodically sends configuration BPDUs from each port without modifying the BPDUs.	• Port A1: {0, 0, 0, Port A1} • Port A2: {0, 0, 0, Port A2}
DeviceB	• Port B1 receives the configuration BPDU {0, 0, 0, Port A1} from Port A1 and finds it superior to its own configuration BPDU {1, 0, 1, Port B1}, so Port B1 updates its configuration BPDU. • Port B2 receives the configuration BPDU {2, 0, 2, Port C2} from Port C2 and finds it inferior to its own configuration BPDU {1, 0, 1, Port B2}, so Port B2 discards the received configuration BPDU.	• Port B1: {0, 0, 0, Port A1} • Port B2: {1, 0, 1, Port B2}
	• DeviceB compares the configuration BPDU on each port and finds that Port B1 has an optimal configuration BPDU. DeviceB selects Port B1 as the root port and retains the configuration BPDU on Port B1. • DeviceB calculates the configuration	• Root port Port B1: {0, 0, 0, Port A1} • Designated port (Port B2): {0, 5, 1, Port B2}

Device	comparison process	Configuration BPDU After Comparison
	BPDU {0, 5, 1, Port B2} for Port B2 based on the configuration BPDU and path cost of the root port, and compares the calculated configuration BPDU with the original configuration BPDU {1, 0, 1, Port B2} on Port B2. The calculated configuration BPDU is superior to the original one, so DeviceB selects Port B2 as the designated port, replaces Port B2's configuration BPDU with the calculated one, and periodically sends configuration BPDUs from Port B2.	
DeviceC	• Port C1 receives the configuration BPDU {0, 0, 0, Port A2} from Port A2 and finds it superior to its own configuration BPDU {0, 0, 0, Port C1}, so Port C1 updates its configuration BPDU. • Port C2 receives the configuration BPDU {1, 0, 1, Port B2} from Port B2 and finds it superior to its own configuration BPDU {1, 0, 1, Port C2}, so Port C2 updates its configuration BPDU.	• Port C1: {0, 0, 0, Port A2} • Port C2: {1, 0, 1, Port B2}
	• DeviceC compares the configuration BPDU on each port and finds that the configuration BPDU on Port C1 is optimal. DeviceC selects Port C1 as the root port and retains the configuration BPDU on Port C1. • DeviceC calculates the configuration BPDU {0, 10, 2, Port C2} for Port C2 based on the configuration BPDU and path cost of the root port, and compares the calculated configuration BPDU with the original configuration BPDU {1, 0, 1, Port B2} on Port C2. The calculated configuration BPDU is superior to the original one, so DeviceC selects Port C2 as the designated port and replaces its	• Root port (Port C1): {0, 0, 0, Port A2} • Designated port (Port C2): {0, 10, 2, Port C2}

Device	comparison process	Configuration BPDU After Comparison
	configuration BPDU with the calculated one.	
	• Port C2 receives the configuration BPDU {0, 5, 1, Port B2} from Port B2 and finds it superior to its own configuration BPDU {0, 10, 2, Port C2}, so Port C2 updates its configuration BPDU. • Port C1 receives the configuration BPDU {0, 0, 0, Port A2} from Port A2 and finds it the same as its own configuration BPDU, so Port C1 discards the received configuration BPDU.	• Port C1: {0, 0, 0, Port A2} • Port C2: {0, 5, 1, Port B2}
	• The root path cost of Port C1 is 10 (root path cost 0 in the received configuration BPDU plus the link patch cost 10), and the root path cost of Port C2 is 9 (root path cost 5 in the received configuration BPDU plus the link patch cost 4). DeviceC finds that Port C2 has a smaller root path cost and therefore considers the configuration BPDU of Port C2 superior to that of Port C1. DeviceC then selects Port C2 as the root port and retains its configuration BPDU. • DeviceC calculates the configuration BPDU {0, 9, 2, Port C1} for Port C1 based on the configuration BPDU and path cost of the root port, and finds the calculated configuration BPDU inferior to the original configuration BPDU {0, 0, 0, Port A2} on Port C1. DeviceC blocks Port C1 and does not update its configuration BPDU. Port C1 no longer forwards data until STP recalculation is triggered, for example, when the link between DeviceB and DeviceC is down.	• Blocked port (Port C1): {0, 0, 0, Port A2} • Root port (Port C2): {0, 5, 1, Port B2}

After the topology is stable, the root bridge still sends the configuration BPDU message at the time interval specified by Hello Timer, while the non-root bridge equipment receives the configuration BPDU message from the root port and forwards it through the designated port. If the received configuration BPDU with higher priority than itself, the non-root bridge device will update the configuration BPDU information stored in its corresponding port according to the information carried in the received configuration BPDU.

6.2.2 RSTP principle description

RSTP improves STP. according to the shortcomings of STP, RSTP deletes three port states, adds two new port roles, and fully decouples port attributes according to states and roles; In addition, RSTP also adds some corresponding enhancement features and protection measures to achieve the stability and rapid convergence of the network.

6.2.2.1 Port role

RSTP has four port roles: root port, designated port, Alternate port and Backup port. The functions of the root port and designated port are the same as those defined in STP. The alternate port and backup port are defined as follows:

- From the perspective of configuration BPDU transmission:
  - An alternate port is blocked after learning a configuration BPDU sent from another bridge.
  - A backup port is blocked after learning a configuration BPDU sent from itself.
- From the perspective of user traffic:
  - An alternate port acts as a backup of the root port and provides an alternate path from the designated bridge to the root bridge.
  - A backup port acts as a backup of the designated port and provides a backup path from the root bridge to the related network segment.

After roles of all RSTP ports are determined, the topology convergence is completed.

6.2.2.2 Port state

RSTP reduces the number of port states to 3. According to whether the port forwards user traffic and learns MAC address:

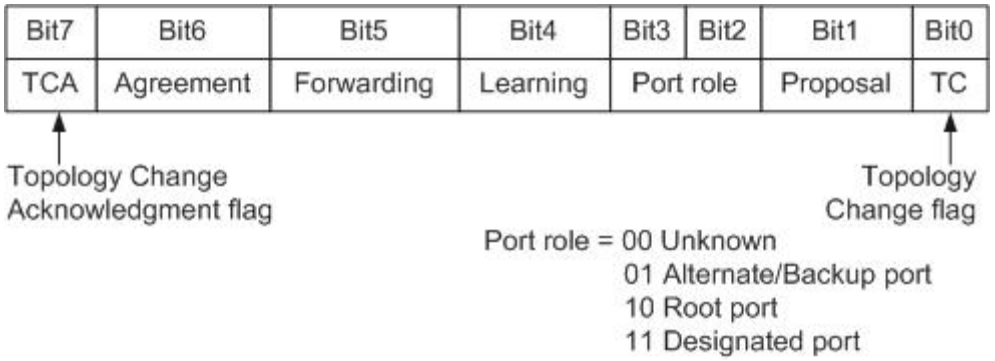
- If the port does not forward user traffic or learn MAC addresses, it is in the Discarding state.
- If the port does not forward user traffic but learns MAC addresses, it is in the Learning state.
- If the port forwards user traffic and learns MAC addresses, it is in the Forwarding state.

6.2.2.3 BPDU format

RSTP retains the basic configuration BPDU format defined in STP with minor changes:

- The value of the Type field is changed from 0 to 2. Devices running STP will discard configuration BPDUs sent from devices running RSTP.
- The Flags field uses the original reserved middle 6 bits, so the changed configuration BPDU is called RST BPDU, as shown in the following figure.

RSTP Flag field format



6.2.2.4 Configure the processing of BPDU

- After the topology is stable, configure the sending mode of BPDU message
  In STP, the root bridge sends configuration BPDUs at Hello timer intervals after the topology becomes stable. Non-root bridges send configuration BPDUs only after they receive configuration BPDUs from upstream devices. This complicates the STP calculation and slows down network convergence. RSTP has improved this, that is, after the topology is stable, whether the non-root bridge device receives the configuration BPDU message from the root bridge or not, the non-root bridge device still sends the configuration BPDU according to the time interval specified by Hello Timer, and this behavior is completely independent by each device.
- BPDU timeout period
  In STP, a device has to wait for a Max Age period before determining a

negotiation failure. In RSTP, a device determines that the negotiation between its port and the upstream device has failed if the port does not receive any configuration BPDUs sent from the upstream device within the timeout interval (Hello Time x 3 x Timer Factor).

- Processing of inferior BPDUs
  When an RSTP port receives an RST BPDU from the upstream designated bridge, the port compares the received RST BPDU with its own RST BPDU.
  - If its RST BPDU is superior to the received one, the port discards the received RST BPDU and immediately responds to the upstream device with its own RST BPDU.
  - After receiving the RST BPDU, the upstream device replaces its RST BPDU with the received RST BPDU. This allows RSTP to rapidly process inferior BPDUs without relying on timers.

  In this manner, RSTP processes inferior BPDUs more rapidly, independent of any timer.

6.2.2.5 Rapid convergence

- Proposal/Agreement mechanism
  In STP, a port that is selected as a designated port needs to wait at least one Forward Delay interval in the Learning state before it enters the Forwarding state. In RSTP, a port that is selected as a designated port enters the Discarding state, and then the proposal/agreement mechanism allows the port to immediately enter the Forwarding state. The proposal/agreement mechanism must be applied on P2P links in full-duplex mode.
- Fast switchover of the root port
  If a root port fails, the best alternate port becomes the root port and enters the Forwarding state. This is because the network segment connected to this alternate port has a designated port connected to the root bridge.
- Edge ports
  In RSTP, a designated port on the network edge is called an edge port. An edge port directly connects to a terminal and does not connect to any other switching devices.
  An edge port does not participate in RSTP calculation. This port can transition from Disabled state to Forwarding state immediately. An edge port becomes a common STP port once it is connected to a switching device and receives a configuration BPDU. The spanning tree needs to be recalculated, which leads to network flapping.

6.2.2.6 Protection functions

RSTP provides the following functions:

- BPDU Guard

  On a switching device, ports directly connected to a user terminal such as a PC or file server are edge ports. Usually, no RST BPDUs are sent to edge ports. If a switching device receives malicious RST BPDUs on an edge port, the switching device automatically sets the edge port to a non-edge port and performs STP calculation. This causes network flapping.

  BPDU protection enables a switching device to set the state of an edge port to error-down if the edge port receives an RST BPDU. In this case, the port remains the edge port, and the switching device sends a notification to the NMS.

- Root protection

  The root bridge on a network may receive superior RST BPDUs due to incorrect configurations or malicious attacks. When this occurs, the root bridge can no longer serve as the root bridge and the network topology will incorrectly change. As a result, traffic may be switched from high-speed links to low-speed links, leading to network congestion.

  If root protection is enabled on a designated port, the port role cannot be changed. When the designated port receives a superior RST BPDU, the port enters the Discarding state and does not forward packets. If the port does not receive any superior RST BPDUs within a specified period (two Forward Delay periods by default), the port automatically enters the Forwarding state.

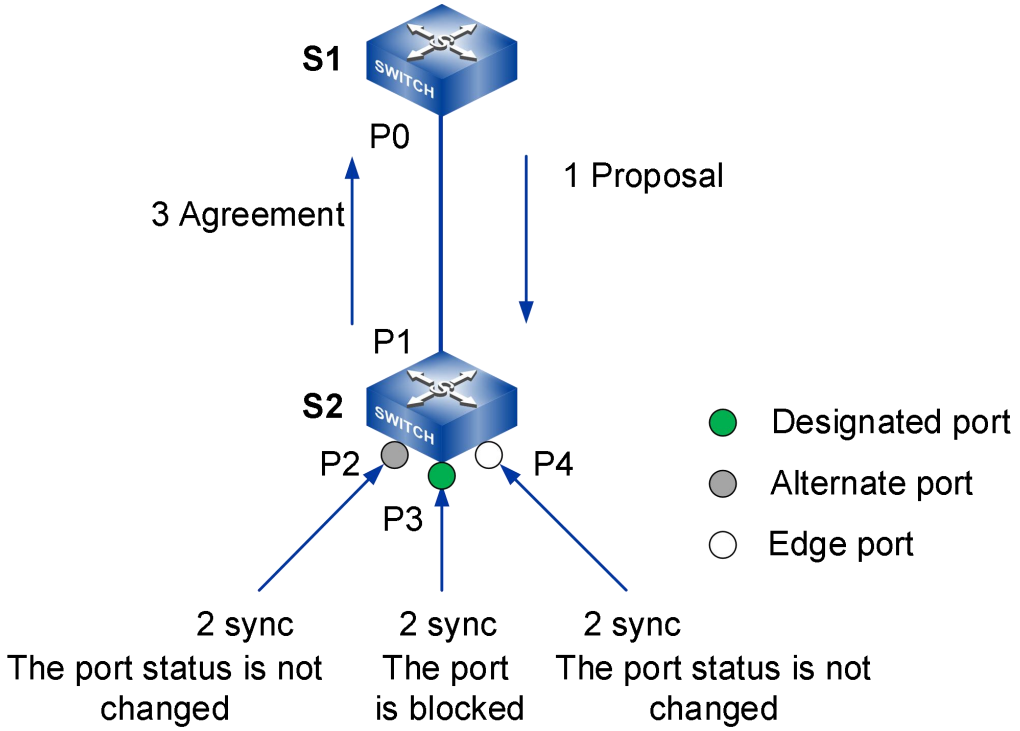
- TC BPDU attack defense

  A switching device deletes its MAC address entries and ARP entries after receiving TC BPDUs. If an attacker sends a large number of malicious TC BPDUs to the switching device within a short period, the device will constantly delete MAC address entries and ARP entries. This increases the load on the switching device and threatens network stability.

  After enabling TC BPDU attack defense on a switching device, you can set the number of TC BPDUs that the device can process within a specified period. If the number of TC BPDUs that the switching device receives within a given time period exceeds the specified threshold, the switching device processes only the specified number of TC BPDUs. After the time period expires, the switching devices process all the excess TC BPDUs together. This function prevents the switching device from frequently deleting MAC entries and ARP entries.

6.2.2.7 Proposal/Agreement mechanism

The Proposal/Agreement mechanism enables a designated port to enter the Forwarding state quickly. As shown, a new link is added between the root bridges S1 and S2. On S2, p2 is an alternate port, p3 is a designated port in Forwarding state, and p4 is an edge port.



The Proposal/Agreement mechanism works as follows:

1 p0 and p1 become designated ports and send RST BPDUs to each other.

2 The RST BPDU sent from p0 is superior to that of p1, so p1 becomes a root port and stops sending RST BPDUs.

3 p0 enters the Discarding state and sets the Proposal and Agreement fields in its RST BPDU to 1.

4 After S2 receives an RST BPDU with the Proposal field set to 1, it sets the sync variable to 1 for all its ports.

5 As p2 has been blocked, its state remains unchanged. p4 is an edge port and does not participate in calculation, so only the non-edge designated port p3 needs to be blocked.

6 After both p2 and p3 enter Discarding state, the synced variable of the port is set, and the synced of the root port p1 is also set, so the response RST BPDU of the Agreement bit is returned to S1. The RST BPDU carries the same information as the BPDU sent by the root bridge just now, except for the Agreement bit (the

Proposal bit is cleared).

7 After S1 receives this RST BPDU, it identifies that the RST BPDU is a response to the proposal that it has sent. Then p0 immediately enters the Forwarding state.

The proposal/agreement process can proceed to downstream devices.

6.2.2.8 RSTP Topology Changes

RSTP considers that the network topology has changed when a non-edge port transitions to the Forwarding state.

When detecting a topology change, RSTP devices react as follows:

- A TC While Timer is started for all non-edge designated ports of this switching device, and the timer value is twice that of Hello Time.
  Within the TC While time, the local device clears MAC address entries learned on all ports.
  At the same time, the non-edge ports send out RST BPDUs with the TC bit set to 1. When the TC While timer expires, the ports stop sending RST BPDUs.
- When other switching devices receive RST BPDUs, they clear MAC address entries learned on all their ports except the ports that receive the RST BPDUs. These switching devices also start a TC While timer on each non-edge designated port and root port and repeat the preceding process.

RST BPDUs are then flooded on the entire network.

6.2.3 MSTP Principle Description

6.2.3.1 Basic Concepts

MSTP divides a switching network into multiple domains, and multiple spanning trees are formed in each domain, which are independent of each other. Each spanning tree is called a Multiple Spanning Tree Instance (MSTI) and each region is called a Multiple Spanning Tree (MST) region. Figure 2 shows an example of an MST region. A spanning tree instance is a collection of multiple VLANs. Binding multiple VLANs to a single MSTI reduces communication costs and resource usage. The topology of each MSTI is calculated independently, and traffic can be balanced among MSTIs. Multiple VLANs with the same topology can be mapped to a single MSTI.The forwarding state of the VLANs for a port is determined by the port state in the MSTI.

MSTP links VLAN with MSTI by setting VLAN mapping table (that is, the correspondence table between VLAN and MSTI). This means that traffic of a VLAN

can be transmitted in only one MSTI.An MSTI, however, can correspond to multiple VLANs.

MST Region

An MST region contains multiple network segments, each of which contains one or more switches. The switches in one MST region all share the following characteristics:

- MSTP-enabled
- Same region name
- Same VLAN-MSTI mappings
- Same MSTP revision level

A local area network can have multiple MST domains, which are physically connected directly or indirectly. Users can divide multiple switching devices into the same MST domain by MSTP configuration command.

VLAN Translation Mapping Table

Each MST region has a VLAN mapping table. The VLAN mapping table maps VLANs to MSTIs.

CST

A Common Spanning Tree (CST) connects all MST regions on a switching network.

The CST is calculated using STP or RSTP, with each MST region being considered as a single node.

IST

An Internal Spanning Tree (IST) resides within an MST region.

An IST is a special MSTI with an MSTI ID of 0.

An IST is a segment of the CIST in an MST region.

SST

A Single Spanning Tree (SST) is formed in either of the following situations:

- A switch running STP or RSTP belongs to only one spanning tree.
- An MST region has only one switch.

CIST

A Common and Internal Spanning Tree (CIST) connects all the switches on a switching network and is calculated using STP or RSTP.

IST and CST of all MST domains constitute a complete spanning tree, namely CIST.

Regional Root

Regional roots are classified into Internal Spanning Tree (IST) and MSTI regional roots.

In MST domain, the nearest switching device in IST spanning tree to CIST Root is IST domain root.

An MST region can contain multiple spanning trees, each of which is called an MSTI. An MSTI regional root is the root of the MSTI. In Figure 2, each MSTI has its own regional root.

Common Root Bridge

In Figure 1, the CIST root is the root bridge of the CIST.

Master Bridge

Master Bridge, also known as IST Master, is the switching equipment closest to the total root in the domain.

If the CIST root is in an MST region, the CIST root is the master bridge of the region.

Port role

The functions of root port, designated port, Alternate port, Backup port and edge port are as defined in RSTP protocol, and the roles of all ports defined in MSTP are shown in the following table.

Port role	Note
Root Port	A root port sends data to a root bridge and is the port closest to the root bridge. Root bridges do not have root ports. Root ports are responsible for sending data to root bridges.
Designated Port	The designated port on a switch forwards BPDUs to a downstream switch.
Alternate port	• Alternate ports provide an alternate path to the root bridge. This path is different from the path through the root port. • An alternate port is blocked from sending BPDUs after a BPDU sent by another bridge is received.
Backup port	• Backup ports provide a backup path to a segment already connected by a designated port. • Backup ports are blocked from sending BPDUs after a BPDU sent by itself is received.
Master port	Master port is the shortest path among all paths connecting MST domain and total root, and it is the port connecting MST domain to total root on switching equipment. BPDUs of an MST region are sent to the CIST root through the master port. Master ports are special regional edge ports, functioning as root ports on CISTs and master ports in instances.

Port role	Note
Regional edge port	A regional edge port is located at the edge of an MST region and connects to another MST region or an SST.
Edge	An edge port is located at the edge of an MST region and does not connect to any switching device. Generally, edge ports are directly connected to terminals. After MSTP is enabled on a port, edge port detection is started automatically. If the port fails to receive BPDU packets within (2 x Hello Timer + 1) seconds, the port is set to an edge port. Otherwise, the port is set to a non-edge port.

MSTP Port States

The port status defined by MSTP is the same as that defined in RSTP protocol, as shown in the following table.

Port state	Note
Forwarding	A port in this state can send and receive BPDUs. It can also forward user traffic.
Learning	This is a transitional state. When a port is in Learning state, it can send and receive BPDUs, but does not forward user traffic. The device creates MAC address entries based on user traffic received on the port but does not forward user traffic through the port. In Learning state, the port can send and receive BPDUs, but cannot forward user traffic.
Discarding	A port in Discarding state can only receive BPDUs.

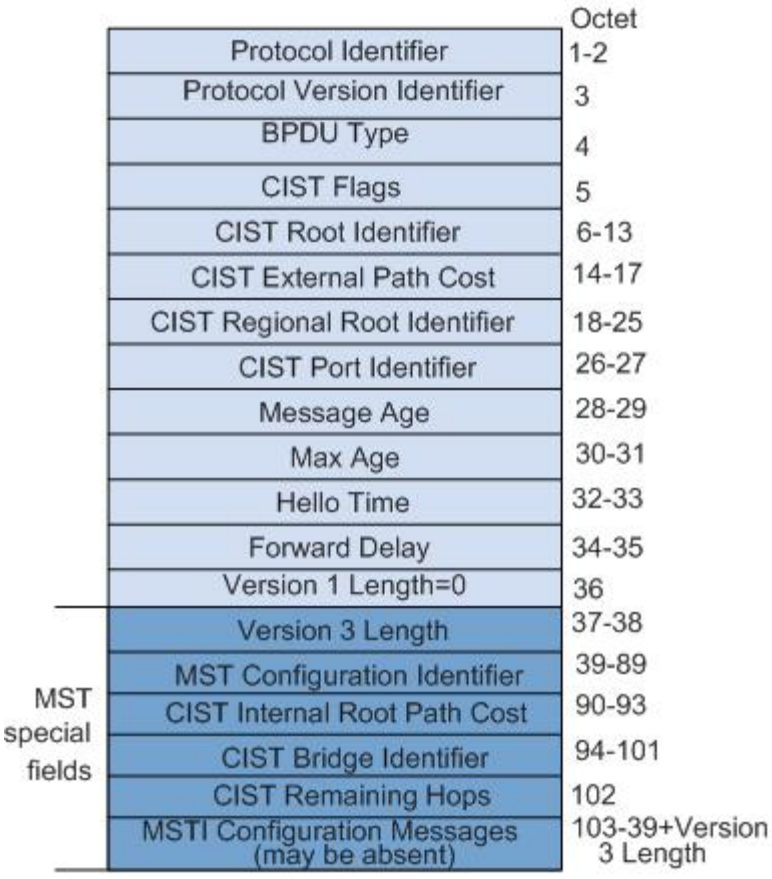
6.2.3.2 MST BPDUs

MSTP calculates spanning trees based on Multiple Spanning Tree Bridge Protocol Data Units (MST BPDUs). Switches on an MSTP network transmit MST BPDUs to calculate spanning tree topologies, maintain network topologies, and communicate topology changes.

The following table shows the difference pairs of configuration BPDU defined in STP, RST BPDU defined in RSTP, MST BPDU defined in MSTP and TCN BPDU.

Version	Type	Name
0	0x00	Configuration BPDU
0	0x80	TCN BPDU
2	0x02	RST BPDU
3	0x02	MST BPDU

The MST BPDU message structure is shown in the following figure.



The first 36 bytes of an MST BPDU are the same as those of an RST BPDU.

Fields from the 37th byte of an MST BPDU are MSTP-specific. The MSTI Configuration Messages field consists of configuration messages of multiple MSTIs.

The main information in MST BPDU is shown in the following table.

Field	Bytes	Note
Protocol Identifier	2	Identifies a protocol.
Protocol Version Identifier	1	Indicates the protocol version identifier: 0:

Field	Bytes	Note
		STP 2: RSTP 3: MSTP
BPDU Type	1	Indicates the BPDU type: • 0x00: Configuration BPDU for STP • 0x80: TCN BPDU (Topology Change Notification BPDU) for STP • 0x02: RST BPDU (Rapid Spanning-Tree BPDU) or MST BPDU (Rapid Spanning-Tree BPDU)
CIST Flags	1	Identifies the CIST.
CIST Root Identifier	8	Indicates the ID of the CIST root switch.
CIST External Path Cost	4	Indicates the total path cost from the MST region where the switch resides to the MST region where the CIST root switch resides. This value is calculated based on link bandwidth.
CIST Regional Root Identifier	8	Indicates the ID of the regional root switch on the CIST. If the root is in this region, the CIST Regional Root Identifier is the same as the CIST Root Identifier.
CIST Port Identifier	2	Indicates the ID of the designated port in the IST.
Message Age	2	Indicates the lifecycle of the BPDU.
Max Age	2	Indicates the maximum lifecycle of the BPDU. If the Max Age timer expires, the link to the root is considered faulty.
Hello Time	2	Indicates the Hello timer value. The default value is 2 seconds.
Forward Delay	2	Indicates the forwarding delay timer. The default value is 15 seconds.
Version 1 Length	1	Indicates the BPDUv1 length, which has a fixed value of 0.

Field	Bytes	Note
Version 3 Length	2	Indicates the BPDUv3 length.
MST Configuration Identifier	51	Indicates the MST configuration identifier, which has four fields.
CIST Internal Root Path Cost	4	Indicates the total path costs from the local port to the IST master. This value is calculated based on link bandwidth.
CIST Bridge Identifier	8	Indicates the ID of the designated switch on the CIST.
CIST Remaining Hops	1	Indicates the number of remaining hops of the BPDU in the CIST.
MSTI Configuration Messages(may be absent)	16	Indicates an MSTI configuration message. Each MSTI configuration message occupies 16 bytes.If there are n MSTIs, MSTI configuration messages are n*16 bytes long.

6.2.3.3 MSTP Topology Calculation

MSTP can divide the entire Layer 2 network into multiple MST regions. The CST is generated through calculation. In an MST region, multiple spanning trees are calculated, each of which is called an MSTI. Among these MSTIs, MSTI 0 is also known as the internal spanning tree (IST). MSTP, like STP, USES configuration messages for spanning tree calculations, except that the configuration messages carry the configuration information of the MSTP on the device.

Vectors

Both MSTI and CIST are calculated according to priority vectors, and these priority vector information are contained in MST BPDU. The exchange devices exchange MST BPDU with each other to generate MSTI and CIST.

Introduction to priority vector

- The priority vectors participating in CIST calculation are:
  Root switching device ID, external path cost, domain root ID, internal path cost, designated switching device ID, designated port ID, and receiving port ID
- The priority vectors participating in MSTI calculation are:

Domain root ID, internal path cost, specified switching device ID, specified port ID, and receiving port ID

The priority of vectors decreases from left to right.

Vector description:

Vector Name	Note
Root ID	Identifies the root switch for the CIST. The root identifier consists of the priority value (16 bits) and MAC address (48 bits). The priority value is the priority of MSTI 0.
External root path cost (ERPC)	Indicates the path cost from a CIST regional root to the root. ERPCs are the same on all switches in an MST region. If the CIST root is in an MST region, all ERPCs in that MST region are set to 0.
Regional root ID	The domain root ID is used to select the domain root in MSTI. Domain root ID = Priority(16bits) +MAC(48bits). The priority value is the priority of MSTI 0.
Internal root path cost (IRPC)	Indicates the path cost from the local bridge to the regional root. The IRPC saved on a regional edge port must be greater than the IRPC saved on a non-regional edge port.
Designated switching device ID	Identifies the nearest upstream bridge on the path from the local bridge to the regional root. If the local bridge is the root or the regional root, this ID is the same as the local bridge ID.
Designated port ID	Identifies the port on the designated switch connected to the root port on the local bridge. Port ID = Priority (4 bits) + port number (12 bits). The priority value must be a multiple of 16.
Receiving port ID	Identifies the port receiving the BPDU. Port ID = Priority(4 bits) + port number (12 bits). The priority value must be a multiple of 16.

Comparative Principle

Compared with the same vector, the vector with the smallest value has the highest priority.

The priority vector comparison principle is as follows.

- First, compare the root switching device IDs.
- If the root switching device ID is the same, compare the external path cost.
- If the external path costs are the same, compare the domain root ID.

- If the domain root ID is still the same, compare the internal path cost.
- If the internal path is still the same, compare the specified switching device ID.
- If the specified switching device ids are still the same, compare the specified port ID.
- If the specified port ids are the same, compare the receiving port ID.

If the configuration message contained in the BPDU received by the port is better than the configuration message stored on the port, the original configuration message stored on the port is replaced by the newly received configuration message. The port simultaneously updates the global configuration message saved by the switching device. Otherwise, the newly received BPDU is discarded.

CIST Calculation

After comparing the vectors, the switch with the highest priority on the entire network is selected as the CIST root. MSTP calculates an IST for each MST region, and calculates a CST to interconnect MST regions. The CST and ISTs form a CIST for the entire network.

MSTI Calculation

In an MST region, MSTP independently calculates an MSTI for each VLAN based on mappings between VLANs and MSTIs. Each spanning tree is calculated independently, and the calculation process is similar to that of STP.

MSTIs have the following characteristics:

- The spanning tree is calculated independently for each MSTI. Spanning trees of MSTIs are independent of each other.
- The spanning tree calculation method of each MSTI is basically the same as STP.
- Spanning trees of MSTIs can have different roots and topologies.
- Each MSTI sends BPDUs in its spanning tree.
- The topology of each MSTI is configured by using commands.
- A port can be configured with different parameters for different MSTIs.
- A port can play different roles or have different status in different MSTIs.

On an MSTP-aware network, a VLAN packet is forwarded along the following paths:

- MSTI in an MST region
- Between the MST domains, it forwards along CST.

6.2.3.4 MSTP Fast Convergence

The P/A mechanism implementation supported by MSTP in common mode is the same as that supported by RSTP.

6.3 Configure STP/RSTP/MSTP

6.3.1 Global Spanning-tree Enablement

【Command】

spanning-tree (enable | disable)

【View】

Configure Mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree enable: this command is used to enable the global spanning tree protocol.

spanning-tree disable: this command is used to disable the spanning tree protocol.

By default, the global spanning tree protocol is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree disable
```

6.3.2 MSTP Instance Configuration

6.3.2.1 Enter MSTP Instance Configuration View

【Command】

spanning-tree mst configuration

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree mst configuration: this command is used to enter the MSTP instance configuration view.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree mst configuration
```

6.3.2.2 Create MSTP Instance

【Command】

instance < instance -id> vlan <vlan-id>
no instance < instance -id> [vlan <vlan-id>]

【View】

MST configuration view

【Default Level】

2: Configuration level

【Parameter】

<instance-id>: represents the number of MSTI in the range 1-4094.

<vlan-id>: VLAN ID value, range is 1-4094.

【Description】

instance: This command is used to create MSTP instance.

no instance: this command is used to delete the MSTP instance.

instance instance_id vlan vlan_id: this command is used to configure the mapping between vlan and MSTP instances. If the Instance does not exist, it will be created first.

By default, all VLANS map to CIST (that is, MSTI 0).

Notice

- Different MSTI cannot be mapped to the same VLAN.
- When adding a VLAN mapping, it is recommended to configure the VLAN first.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree mst configuration
Switch(config-mst)#instance 1 vlan 2
```

6.3.2.3 MSTP Revision Level

【Command】

```
revision <level>
```

【View】

MST configuration view

【Default Level】

2: Configuration level

【Parameter】

<level>: revision level, range 0-65535.

【Description】

revision: this command is used to configure the MSTP revision level for the MST domain.

By default, the MSTP revision level for the MST domain is 0.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree mst configuration
Switch(config-mst)#revision 1
```

6.3.2.4 MST Domain Name

【Command】

```
region <name>
no region <name>
```

【View】

MST configuration view

【Default Level】

2: Configuration level

【Parameter】

<name> : the domain name of the MST domain.

【Description】

region <name>: this command is used to configure the domain name for the MST domain.

By Default, the MST domain name is Default.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree mst configuration
Switch(config-mst)#region test
```

6.3.3 Bridge Configuration

6.3.3.1 Device Priority

【Command】

spanning-tree [instance <instance_id>] priority <priority>
no spanning-tree [instance <instance_id>] priority

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<priority>: device priority (multiple of 4096), <0-61440>

【Description】

spanning-tree priority: this command is used to configure the device priority of CIST.

no spanning-tree priority: this command is used to restore the device priority of CIST to the default value.

spanning-tree instance priority: this command is used to configure the device priority of the MSTI.

no spanning-tree instance priority: this command is used to restore the device priority of MSTI to the default value.

By default, the device priority is 32768.

CIST refers to spanning tree instance 0, and MSTI refers to creating spanning tree instance, with instance range 1-4094.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree priority 4096
```

6.3.3.2 Spanning-tree Protocol Version

【Command】

```
spanning-tree force-version <version>
no spanning-tree force-version
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<version>: spanning tree protocol version number, range is 0- 3, 0 is STP compatibility mode, 2 is RSTP mode, 3 is MSTP mode, 1 is unsupported.

【Description】

spanning-tree force-version: this command is used to configure the version of the spanning tree protocol.

no spanning-tree force-version: this command restores the version of the spanning tree protocol to the MSTP protocol.

By default, the working mode of MSTP is MSTP mode.

MSTP and RSTP can recognize each other's protocol messages and are compatible with each other. STP cannot recognize MSTP messages, in order to realize mixed networking with STP equipment and complete compatibility with RSTP, there are three operating modes have been set: STP compatibility mode, RSTP mode and MSTP mode.

- In STP compatibility mode, each port of the device will send out STP BPDU messages.
- In the RSTP mode, each port of the device will send out RSTP BPDU messages. When it is found to be connected with the device running STP, the port will automatically switch to the STP compatibility mode.
- In MSTP mode, each port of the device will send out MSTP BPDU messages. When it is found that it is connected with the device running STP, the port will

automatically switch to STP compatibility mode.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree force-version 2
```

6.3.3.3 Spanning Tree Timer Parameter

【Command】

```
spanning-tree (hello-time | forward-time | max-age) <seconds>
no spanning-tree hello-time
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<seconds>: hello-time range <1-10>, forward-time range <4-30>, max-age range <6-40>, all in seconds

【Description】

spanning-tree hello-time: the command is used to configure the hello-time.

no spanning-tree hello-time: this command is used to restore the hello-time to default value.

By default, hello-time is 2 seconds, forward-time is 15 seconds, and max-age is 20 seconds.

The values of the three time parameters of the root bridge hello-time, forward-time and max-age should meet the following formula, otherwise will cause the network oscillation frequently:

2 ×（forward-time - 1）≥ max-age

max-age ≥ 2 ×（hello-time + 1）

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree hello-time 3
```

6.3.3.4 The Maximum Hop of Spanning-tree

【Command】

```
spanning-tree max-hops <hops>
no spanning-tree max-hops
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<hops> : the maximum hops of MST domain are in the range <1-40>.

【Description】

spanning-tree max-hops： this command is used to configure the maximum number of hops of the MST domain.

no-spanning -tree max-hops: this command restores max-hops as the default. By default, the max-hops is 20.

Starting from the root bridge of spanning tree in MST domain, every time configuration message in domain (namely BPDU message) is forwarded by a device, the hop number is reduced by 1; Devices discard configuration messages with 0 hops received, preventing devices outside the maximum hops from participating in the spanning tree calculation, limiting the size of the MST domain.

If the current device becomes the root bridge of CIST in MST domain or the root bridge of MSTI, the maximum hop number of this device configuration will become the network diameter of this spanning tree, limiting the scale of this spanning tree in the current MST domain. Devices that do not generate root Bridges in the MST domain will use the maximum number of hops set by the root bridge.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree max-hops 24
```

6.3.3.5 The Rate that the Spanning Tree Sends a BPDU

【Command】

```
spanning-tree transmit-holdcount <count>
no spanning-tree transmit-holdcount
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<count> : maximum number of message sent per second, range is <1-10>.

【Description】

spanning-tree transmit-holdcount: this command is used to configure the maximum rate of sending the BPDU of the port.

no spanning-tree transmit-holdcount: this command is used to restore the maximum rate of sending the BPDU of the port to default value.

By default, transmit-holdcount is 6.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree transmit-holdcount 10
```

6.3.3.6 Compatible with Cisco MSTP Mode

【Command】

spanning-tree cisco-interoperability (enable | disable)

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree cisco-interoperability enable: this command is used to enable compatibility with the cisco MSTP pattern.

By default, cisco MSTP mode is not compatible.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree cisco-interoperability enable
```

6.3.4 Port Configuration

6.3.4.1 Global Edge Port BPDU Filtering

【Command】

```
spanning-tree portfast bpdu-filter
no spanning-tree portfast bpdu-filter
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

`spanning-tree portfast bpdu-filter`: this command is used to enable the global portfast bpdu-filter function.

`no-spanning -tree portfast bpdu-filter`: this command disables the global portfast bpdu-filter function.

By default, global portfast bpdu-filter is disabled.

The portfast features (bpdu-filter and bpdu-guard) both need to be valid on the portfast port (see related command spanning-tree portfast). The portfast feature can be enabled in configure mode or under the port (see the relative commands under the port). There are only enable and disable two state in configure mode, while there are enable, disable and default three states under the port. When the configured portfast feature of the port is default, the actual running portfast feature of the port will be the same as the portfast feature in configure mode. When the portfast feature configured of the port is enable or disable, the actual running portfast feature of the port will be consistent with the portfast feature configured on the port. Use the show spanning-tree interface command to view relative details.

By default, the global portfast bpdu-filter function is disabled.

The portfast function is mainly used to connect devices such as terminals or servers, which needs fast convergence ports, similar to edgeport. Portfast must be enabled if the port needs to use the portfast feature.

Bpdu-filter function is used for portfast port, after enabling it, the port will not send and receive bpdu messages.

bpdu-guard function is used for portfast port. Once enabled, when the port receives bpdu message, the port will change into error-disable state (shutdown), which can be restored by errdisable-timeout function.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree portfast bpdu-filter

*Switch#show spanning-tree interface ge6
% Default: Bridge up - Spanning Tree Enabled  - topology change
detected
% Default: CIST Root Path Cost 0 - CIST Root Port 0 -  CIST Bridge
Priority 32768
% Default: Forward Delay 15 - Hello Time 2 - Max Age 20 - Transmit
Hold Count 6 - Max-hops 20
% Default: CIST Root Id 800000b25f5f0003
% Default: CIST Reg Root Id 800000b25f5f0003
% Default: CIST Bridge Id 800000b25f5f0003
% 0: 1 topology change(s)  - last topology change Thu Jan  1 08:00:29
1970

% Default: portfast bpdu-filter disabled
% Default: portfast bpdu-guard enabled
% Default: portfast errdisable timeout enabled
% Default: portfast errdisable timeout interval 300 sec
%   ge6: Port Number 910 - Ifindex 5006 - Port Id 838e - Role Disabled
- State Discarding
%   ge6: Designated External Path Cost 0 -Internal Path Cost 0
%   ge6: Configured Path Cost 20000  - Add type Explicit ref count
1
%   ge6: Designated Port Id 0 - CIST Priority 128  -
%   ge6: Message Age 0 - Max Age 0
%   ge6: CIST Hello Time 0 - Forward Delay 0
%   ge6: CIST Forward Timer 0 - Msg Age Timer 0 - Hello Timer 0 -
topo change timer 0
%   ge6: forward-transitions 0
%   ge6: Version Multiple Spanning Tree Protocol - Received None
- Send MSTP
%   ge6: No portfast configured - Current  portfast off
```

```
%    ge6: portfast bpdu-guard default - Current portfast
bpdu-guard on
%    ge6: portfast bpdu-filter default - Current portfast
bpdu-filter off
%  ge6: no root guard configured    - Current root guard off
%     ge6: Configured Link Type point-to-point - Current
point-to-point
%  ge6: No auto-edge configured - Current port Auto Edge off
```

6.3.4.2 Global Edge Port BPDU Protection

【Command】

spanning-tree portfast bpdu-guard
no spanning-tree portfast bpdu-guard

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree portfast bpdu-guard: this command is used to enable the global portfast bpdu-guard function.
no spanning-tree portfast bpdu-guard: this command is used to disable the global portfast bpdu-guard function.
By default, global portfast bpdu-guard is disabled.

The portfast features (bpdu-filter and bpdu-guard) both need to be valid on the portfast port (see related command spanning-tree portfast). The portfast feature can be enabled in configure mode or under the port (see the relative commands under the port). There are only enable and disable two state in configure mode, while there are enable, disable and default three states under the port. When the configured portfast feature of the port is default, the actual running portfast feature of the port will be the same as the portfast feature in configure mode. When the portfast feature configured of the port is enable or disable, the actual running portfast feature of the port will be consistent with the portfast feature configured on the port. Use the show spanning-tree interface command to view relative details.

By default, the global portfast bpdu-filter function is disabled.

The portfast function is mainly used to connect devices such as terminals or servers, which need fast convergence ports, similar to edgeport. Portfast must be enabled if the port needs to use the portfast feature.

Bpdu-filter function is used for portfast port, after enabling it, the port will not send and receive bpdu messages.

Bpdu-guard function is used for portfast port. Once enabled, when the port receives bpdu message, the port will change into error-disable state (shutdown), which can be restored by errdisable-timeout function.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree portfast bpdu-guard
```

6.3.4.3 Port error-disable Timeout Recover

【Command】

spanning-tree errdisable-timeout (enable | disable)

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree errdisable-timeout enable: this command is used to configure the error-disable timeout recovery function of the port.

spanning-tree errdisable-timeout disable: this command is used to disable the error-disable timeout recovery function of the port.

By default, the port error-disable timeout recovery function is enabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree errdisable-timeout enable
```

6.3.4.4 Port error-disable Recovery Interval

【Command】

spanning-tree errdisable-timeout interval <seconds>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<seconds>: the recover interval of error-disable, in the range <10-1000000>.

【Description】

spanning-tree errdisable-timeout interval: this command is used to configure the time interval for recovery after error-disable of the port .

no spanning-tree errdisable-timeout interval: this command is used to restore the time interval for recovery after error-disable of the port to default.

errdisable-timeout interval is 300 seconds by default.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#spanning-tree errdisable-timeout interval 400
```

6.3.4.5 Edge Port Enabled

【Command】

spanning-tree portfast

no spanning-tree portfast

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree portfast: this command is used to enable the portfast function of the port.

no spanning-tree portfast: this command is used to disable the portfast function.

By default, port portfast is disabled.

The portfast function is mainly used to connect terminals or servers and other devices, requiring fast convergence of the port. Portfast must be enabled if the port needs to use the portfast feature.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree portfast
```

6.3.4.6 BPDU Filter of Edge Port

【Command】

spanning-tree portfast bpdu-filter (enable | disable | default)

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree portfast bpdu-filter: this command is used to configure the mode of the portfast bpdu-filter feature under the port, enable, disable, default respectively.

By default, the port portfast bpdu-filter is in default.

The portfast features (bpdu-filter and bpdu-guard) both need to be valid on the portfast port (see related command spanning-tree portfast). The portfast feature can be enabled in configure mode or under the port (see the relative commands under the port). There are only enable and disable two state in configure mode, while there are enable, disable and default three states under the port. When the configured portfast

feature of the port is default, the actual running portfast feature of the port will be the same as the portfast feature in configure mode. When the portfast feature configured of the port is enable or disable, the actual running portfast feature of the port will be consistent with the portfast feature configured on the port. Use the show spanning-tree interface command to view relative details.

By default, the global portfast bpdu-filter function is disabled.

The portfast function is mainly used to connect devices such as terminals or servers, which needs fast convergence ports, similar to edgeport. Portfast must be enabled if the port needs to use the portfast feature.

Bpdu-filter function is used for portfast port, after enabling it, the port will not send and receive bpdu messages.

bpdu-guard function is used for portfast port. Once enabled, when the port receives bpdu message, the port will change into error-disable state (shutdown), which can be restored by errdisable-timeout function.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree portfast bpdu-filter enable
```

6.3.4.7 Edge Port BPDU Guard

【Command】

spanning-tree portfast bpdu-guard (enable | disable | default)

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree portfast bpdu-guard: this command is used to configure the modes of the portfast bpdu-guard feature of the port, they are enable, disable, default.

By default, the port portfast bpdu-guard is in default.

The portfast features (bpdu-filter and bpdu-guard) both need to be valid on the portfast port (see related command spanning-tree portfast). The portfast feature can be enabled in configure mode or under the port (see the relative commands under the port). There are only enable and disable two state in configure mode, while there are enable, disable and default three states under the port. When the configured portfast feature of the port is default, the actual running portfast feature of the port will be the same as the portfast feature in configure mode. When the portfast feature configured of the port is enable or disable, the actual running portfast feature of the port will be consistent with the portfast feature configured on the port. Use the show spanning-tree interface command to view relative details.

By default, the global portfast bpdu-filter function is disabled.

The portfast function is mainly used to connect devices such as terminals or servers, which need fast convergence ports, similar to edgeport. Portfast must be enabled if the port needs to use the portfast feature.

Bpdu-filter function is used for portfast port, after enabling it, the port will not send and receive bpdu messages.

bpdu-guard function is used for portfast port. Once enabled, when the port receives bpdu message, the port will change into error-disable state (shutdown), which can be restored by errdisable-timeout function.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree portfast bpdu-guard enable
```

6.3.4.8 Automatical Switching Edge Port

【Command】

```
spanning-tree autoedge
no spanning-tree autoedge
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree autoedge: this command is used to configure ports to automatically switch to edge ports.

no spanning-tree autoedge: this command conjures ports that cannot be automatically switched to non-edge ports.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree autoedge
```

6.3.4.9 Root Port Protection

【Command】

spanning-tree guard root
no spanning-tree guard root

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree guard root: this command is used to configure the root port protection function.

no spanning-tree guard root: this command concatenates the port to not enable root port protection.

By default, root port guard is not enabled.

guard root is a mandatory root protection that stops accidental (or illegal) switches becoming root Bridges in the network. when the guard root port (designated ports) is opened and receives better BPDU packets, the port will enter a Listening (STP) or discarding state (RSTP, MSTP).

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
```

```
Switch(config-ge8)#spanning-tree guard root
```

6.3.4.10 Port Spanning-tree Enablement

【Command】

```
spanning-tree (enable | disable)
```

【View】

Interface port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

spanning-tree enable: this command is used to enable the spanning tree function of the port.

spanning-tree disable: this command is used to disable the spanning tree function of the port.

By default, the port spanning tree function is enabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree enable
```

6.3.4.11 Port Hello-time

【Command】

```
spanning-tree hello-time <seconds>
no spanning-tree hello-time <seconds>
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<seconds> ： hello-time，range is 1-10.

【Description】

spanning-tree hello-time: this command is used to configure the hello-time of the port.

no spanning-tree hello-time: this command is used to restore the hello-time of the port to its default value.

By default, the hello-time of the port is 2.

It is better to use global configuration commands for hello-time.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree hello-time 3
```

6.3.4.12 Port Connection Type

【Command】

spanning-tree link-type (auto | point-to-point | shared)

no spanning-tree link-type

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

Auto: automatically determines the connection type, point-to-point or shared, based on duplex mode.

Point-to-point: specifies the port type as point-to-point.

Shared: specifies the port type as shared.

【Description】

spanning-tree link-type: this command is used to modify the link type of the port.

no spanning-tree link-type: this command is used to restore the link type of the port to the default value.

By default, the link type of the port is auto.

【Instance】

```
Switch> enable
Switch#configure terminal
```

```
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree link-type point-to-point
```

6.3.5 Instance Port Configuration

6.3.5.1 Port Priority

【Command】

```
spanning-tree [instance <instance_id>] priority <priority>
no spanning-tree instance <instance_id> priority
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

<instance-id> : the number of MSTI in the range 0-4094.

<priority> : port priority, the range is 0-240.

【Description】

spanning-tree priority: this command is used to configure the port priority.

no spanning-tree priority: this command is used to restore the port priority to the default.

By default, the port priority is 128.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree priority 32
```

6.3.5.2 Cost

【Command】

```
spanning-tree [instance <instance_id>] path-cost <cost>
no spanning-tree instance <instance_id> path-cost
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

<instance-id> : the number of MSTI in the range 0-4094.

<cost> : means the cost value of the port, ranging from 1-200000000.

【Description】

spanning-tree path-cost: this command is used to configure the port cost.

no spanning-tree path-cost: this command is used to restore the port cost as the default.

By default, the port cost is 20000000.

When the port cost is the default value, the actual cost of link up port is converted according to the port rate, the rate of 10M corresponds to the cost of 2000000, and 100M corresponds to the cost of 200000.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree path-cost 1000
```

6.3.5.3 Port Restricted Election

【Command】

```
spanning-tree [instance <instance_id>] restricted-role
no spanning-tree instance <instance_id> restricted-role
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

<instance-id> : the number of MSTI in the range 0-4094.

【Description】

spanning-tree restricted-role: the command is used to configure ports to restrict elections so that ports cannot be elected as root ports.

no spanning-tree restricted-role: the command is used to cancel port restricted elections.

By default, the port does not restrict elections.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree restricted-role
```

6.3.5.4 Port Restriction TC

【Command】

spanning-tree [instance <instance_id>] restricted-tcn
no spanning-tree instance <instance_id> restricted-tcn

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

<instance-id> : the number of MSTI in the range 0-4094.

【Description】

spanning-tree restricted-tcn: the command is used to configure port restriction processing for receiving TC bits in BPDU message.

no spanning-tree restricted-tcn: the command is used to cancel the port restriction processing of the TC bit in the received BPDU message.

By default, no restriction on the port.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#spanning-tree restricted-tcn
```

6.3.6 Display Spanning Tree Information

6.3.6.1 Display Spanning-tree Detail Information

【Command】

show spanning-tree (*interface IFNAME* |)

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

Interface IFNAME: displays status information of the specified port.

【Description】

show spanning-tree: this command is used to display the details of spanning-tree.

【Instance】

```
Switch#show spanning-tree
% Default: Bridge up - Spanning Tree Disabled
% Default: CIST Root Path Cost 0 - CIST Root Port 0 -  CIST Bridge
Priority 32768
% Default: Forward Delay 15 - Hello Time 2 - Max Age 20 - Transmit
Hold Count 6 - Max-hops 20
% Default: CIST Root Id 800000226f0100a3
% Default: CIST Reg Root Id 800000226f0100a3
% Default: CIST Bridge Id 800000226f0100a3
% 0: 0 topology change(s)  - last topology change Thu Jan  1 08:00:00
1970

% Default: portfast bpdu-filter disabled
% Default: portfast bpdu-guard disabled
% Default: portfast errdisable timeout disabled
% Default: portfast errdisable timeout interval 300 sec
%   ge1: Port Number 1 - Ifindex 5001 - Port Id 8001 - Role Disabled
- State Discarding
%   ge1: Link down - Spanning Tree Disabled
%   ge1: Designated External Path Cost 0 -Internal Path Cost 0
%   ge1: Configured Path Cost 20000000  - Add type Explicit ref count
1
%   ge1: Designated Port Id 0 - CIST Priority 128  -
%   ge1: Message Age 0 - Max Age 0
%   ge1: CIST Hello Time 0 - Forward Delay 0
%   ge1: CIST Forward Timer 0 - Msg Age Timer 0 - Hello Timer 0 -
topo change timer 0
%   ge1: forward-transitions 0
%   ge1: Version MSTP - Received None - Send MSTP
%   ge1: Auto edge - On
```

```
%   ge1: Portfast - Off
%   ge1: Edge port - False
%   ge1: Bpdu Guard  - Disabled (Config - default)
%   ge1: Bpdu filter - Disabled (Config - default)
%   ge1: Link Type - point-to-point (Config - auto)
%   ge1: Root Guard - Off
```

6.3.6.2 Display the Basic Information of the Spanning Tree

【Command】

show spanning-tree (*instance <instance-id>*|) brief

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

<instance-id> : the number of MSTI in the range 0-4094.

【Description】

show spanning-tree brief: the command is used to display the basic information of spanning-tree.

【Instance】

```
Switch#show spanning-tree brief
MST  Port           Role        State
0    ge10           Designated   Forwarding
```

6.3.6.3 Display MSTP Configuration Information

【Command】

show spanning-tree mst config

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show spanning-tree mst config: display MSTP configuration information.

【Instance】

```
Switch#show spanning-tree mst config
%
%  MSTP Configuration Information for bridge 0 :
%--------------------------------------------------------
%  Format Id     : 0
%  Name          : Default
%  Revision Level : 0
%  Digest        : 0xAC36177F50283CD4B83821D8AB26DE62
%--------------------------------------------------------
```

6.3.6.4 Display MSTP Detailed Information

【Command】

show spanning-tree mst （detail | interface IFNAME）

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: interface name

【Description】

show spanning-tree mst detail: display the MSTP details of the device.

show spanning-tree mst interface IFNAME: display the MSTP details of the designated port.

【Instance】

```
Switch#show spanning-tree mst interface ge1
% Default: Bridge up - Spanning Tree Disabled
% Default: CIST Root Path Cost 0 - CIST Root Port 0 -  CIST Bridge
Priority 32768
% Default: Forward Delay 15 - Hello Time 2 - Max Age 20 - Transmit
Hold Count 6 - Max-hops 20
% Default: CIST Root Id 800000226f01cca9
% Default: CIST Reg Root Id 800000226f01cca9
% Default: CIST Bridge Id 800000226f01cca9
```

```
% 0: 0 topology change(s)  - last topology change Thu Jan  1 08:00:00
1970

% Default: portfast bpdu-filter disabled
% Default: portfast bpdu-guard disabled
% Default: portfast errdisable timeout disabled
% Default: portfast errdisable timeout interval 300 sec
%
%   Instance       VLAN
%   0:             1
```

6.3.6.5 Display MSTI Instance Information

【Command】

show spanning-tree mst instance <instance-id> (interface IFNAME |)

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

<instance-id> : the number of MSTI in the range 1-4094.

IFNAME: interface name

【Description】

show spanning-tree mst instance <instance-id>: display the information of the designated MSTP instance.

show spanning-tree mst instance <instance-id> interface IFNAME: display the MSTP instance information of the designated port.

【Instance】

```
Switch#show spanning-tree mst instance 1 interface ge1
% 0: MSTI Root Path Cost 0 - MSTI Root Port 0 - MSTI Bridge Priority
8192
% 0: MSTI Root Id 200100226f01cca9
% 0: MSTI Bridge Id 200100226f01cca9
%   ge1: Port Number 1 - Ifindex 5001 - Port Id 8001 - Role Disabled
- State Discarding
%   ge1: Designated Internal Path Cost 0  - Designated Port Id 0
%   ge1: Configured Internal Path Cost 20000000
```

```
%   ge1: Configured CST External Path cost 20000000
%   ge1: CST Priority 128  - MSTI Priority 128
%   ge1: Designated Root 000000226f01cca9
%   ge1: Designated Bridge 000000226f01cca9
%   ge1: Message Age 0 - Max Age 0
%   ge1: Hello Time 0 - Forward Delay 0
%   ge1: Forward Timer 0 - Msg Age Timer 0 - Hello Timer 0
%
```

6.3.6.6 Display MSTI Instance Information of Traffic Engineering

【Command】

show spanning-tree mst instance te-msti

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show spanning-tree mst instance te-msti： display the MSTI instance information of traffic engineering.

【Instance】

```
Switch#show spanning-tree mst instance te-msti
% 0: MSTI Root Path Cost 0 - MSTI Root Port 0 - MSTI Bridge Priority
36864
% 0: MSTI Root Id 900f000000000000
% 0: MSTI Bridge Id 900f000000000000
```

6.3.6.7 Display MSTI Instance VLAN Information

【Command】

show spanning-tree vlan range-index

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show spanning-tree vlan range-index: the command is used to display the corresponding information of MSTI instance and the VLAN.

【Instance】

```
Switch#show spanning-tree vlan range-index
%   Instance  VLAN     RangeIdx
%   1         1        1
```

7 ERPS Configuration

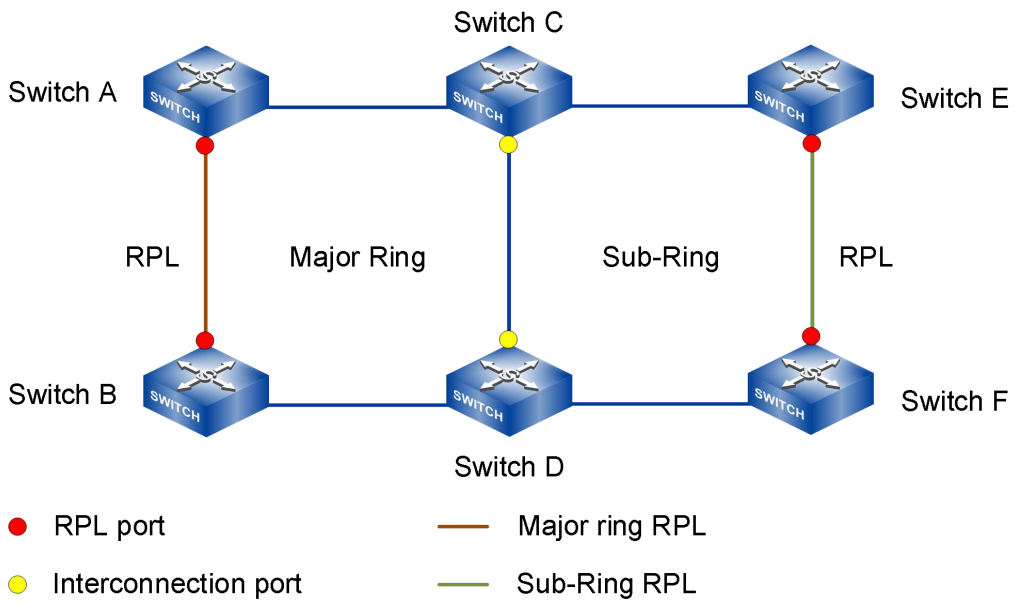
7.1 ERPS Overview

Ethernet Ring Protection Switching (ERPS) is the Ethernet Ring Network Link Layer Technology with high reliability and stability. It can prevent the broadcast storm caused by data loop when the Ethernet ring is intact. ERPS is a protocol defined by the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) to eliminate loops at layer 2. Because the standard number is ITU-T G.8032/Y1344, ERPS is also called G.8032. ERPS defines Ring Auto Protection Switching (RAPS) Protocol Message and protection switching mechanisms. It can prevent the broadcast storm caused by data loop when the Ethernet ring is intact. When the Ethernet ring link failure occurs, it has high convergence speed that can rapidly recover the communication path between each node in the ring network.

7.2 Principle Description

7.2.1 Basic ERPS Concepts

ERPS eliminates loops at the link layer of an Ethernet network. It uses ERPS rings as its basic unit. There are several nodes in an ERPS ring. ERPS blocks the RPL (Ring Protection Link) owner port and controls common ports to switch the port status between Forwarding and Discarding and eliminate loops. ERPS uses the control VLAN, data VLAN, and Ethernet Ring Protection (ERP) instance to better realize ERPS function.



7.2.1.1 ERPS Ring

An ERPS ring consists of interconnected Layer 2 switching devices configured with the same control VLAN.

ERPS ring could be divided into main ring and subring. By default, an ERPS ring is a major ring. The major ring is a closed ring, whereas a sub-ring is a non-closed ring. The major ring and sub-ring are configured using commands. On the network shown in the figure above, SwitchA through SwitchD constitute a major ring, and SwitchC through SwitchF constitute a sub-ring. A node refers to a Layer 2 switching device added to an ERPS ring. A maximum of two ports on each node can be added to the same ERPS ring. SwitchA through SwitchD in the figure above are nodes in an ERPS major ring. The main node is in charge of blocking and opening ports on this node, preventing loops from forming.

Only ERPSv2 supports sub-rings.

7.2.1.2 Port Role

Each device in ERPS ring is called a node. The node role is decided by user configuration, they are divided into following types:

- RPL-Owner: owner node is responsible for blocking and unblocking the port in RPL of the node to prevent loop forming and conduct link switching.
- RPL-Neighbor: neighbor node is connected to Owner node on RPL. Cooperating to the Owner node, it blocks and unblocks the ports on RPL of the node and conduct link switching.
- Interconnection: interconnected node is the node to connect multiple rings in the

multi-loop model, it belongs to the subring, and the primary ring has no interconnected node. In the link protocol packet upload mode between the two subring interconnected nodes, the subring protocol packet ends in the interconnected node, but the data packet won't end.

- Other: normal node. Normal node is responsible for receiving and forwarding the protocol packet and data packet in the link.

7.2.1.3 Port Status

On an ERPS ring, an ERPS-enabled port has two statuses:

- Forwarding: forwards user traffic and sends and receives RAPS PDUs.
- Discarding: only sends and receives RAPS PDUs.

7.2.1.4 Control VLAN

A control VLAN is configured in an ERPS ring to transmit RAPS PDUs.

Each ERPS ring must be configured with a control VLAN. After a port is added to an ERPS ring configured with a control VLAN, the port is added to the control VLAN automatically.

Different ERPS rings must use different control VLANs.

7.2.1.5 Data VLAN

Unlike control VLANs, data VLANs are used to transmit data packets.

7.2.1.6 ERP Instance

On a Layer 2 device running ERPS, the VLAN in which RAPS PDUs and data packets are transmitted must be mapped to an Ethernet Ring Protection (ERP) instance so that ERPS forwards or blocks the packets based on configured rules. If the mapping is not configured, the preceding packets may cause broadcast storms on the ring network. As a result, the network becomes unavailable. One ring in ERPS networking can support multiple instances, and each instance is a logical ring. Each instance has its own protocol channel, data channel and master node. As an independent protocol entity, each instance maintains its own state and data.

7.2.1.7 Timer

In ERPS protocol, timers used mainly include Guard Timer, WTR (Wait to Restore) Timer, Hold Timer and WTB (Wait to Block) Timer.

- Guard timer
  Device involved in link failure or node failure sends NR(No Request) RAPS message to other device after failure recovery or clearing operation, and starts Guard Timer at the same time, and does not process NR RAPS message before the timer expires, in order to prevent receiving expired NR RAPS message. Before the Guard timer expires, the device does not process any RAPS (NR) messages to avoid receiving out-of-date RAPS (NR) messages. After the Guard timer expires, if the device still receives an RAPS (NR) message, the local port enters the Forwarding state.
- WTR timer
  If an RPL owner port is unblocked due to a link or node fault, the involved port may not go Up immediately after the link or node recovers. Blocking the RPL owner port may cause network flapping. Blocking the RPL owner port may cause network flapping. To prevent this problem, the node where the RPL owner port resides starts the wait to restore (WTR) timer after receiving an RAPS (NR) message. The WTR Timer will be turned off if SF(Signal Fail) RAPS messages are received from other ports before the timer expires. If the node does not receive any RAPS (SF) message before the timer expires, it blocks the RPL owner port when the timer expires and sends NR&RB (RPL Block) and RAPS message. After receiving this RAPS (NR, RB) message, the nodes set their recovered ports on the ring to the Forwarding state.
- Hold Timer
  On Layer 2 networks running ERPS, there may be different requirements for protection switching. For example, on a network where multi-layer services are provided, after a server fails, users may require a period of time to rectify the server fault so that clients do not detect the fault. Users can set the Hold timer. If the fault occurs, the fault is not immediately sent to ERPS until the Holdoff timer expires and the fault is still not recovered.
- WTB Timer
  The wait to block (WTB) timer starts when Forced Switch (FS) or Manual Switch (MS) is performed. Because multiple nodes on an ERPS ring may be in FS or MS state, the clear operation takes effect only after the WTB timer expires. This prevents the RPL owner port from being blocked immediately.

7.2.1.8 Revertive and Non-revertive Switching

After link faults in an ERPS ring are rectified, re-blocking the RPL owner port depends on the switching mode:

- In revertive mode, if the failed link recovers, the RPL owner port will be blocked again after waiting for WTR time. Blocked links are switched back to RPL.
- In non-revertive switching, the WTR timer is not started, and the original faulty link is still blocked.

7.2.1.9 Port Blocking Modes

Because the Ring Protection Link (RPL) may have high bandwidth, you can block the low-bandwidth link so that user traffic can be transmitted on the RPL. ERPSv2 supports both Forced Switch (FS) and Manual Switch (MS) modes for blocking an ERPS port:

- FS: forcibly blocks a port immediately after FS is configured, irrespective of whether link failures have occurred.
- MS: blocks a port on which MS is configured when the ERPS ring is in Idle or Pending state.

In addition to FS and MS operations, ERPS also supports the clear operation. The clear operation has the following functions:

- Clears an existing FS or MS operation.
- Triggers revertive switching before the WTR or WTB timer expires in the case of revertive switching operations.
- Triggers revertive switching in the case of non-revertive switching operations.

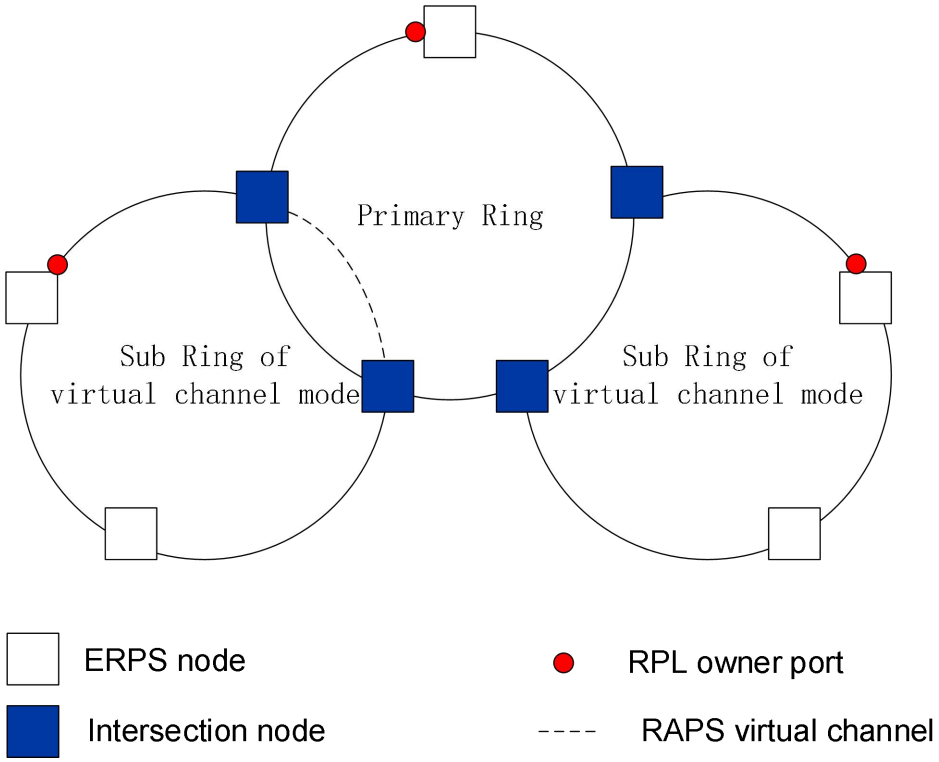
Only ERPSv2 supports port blocking modes.

7.2.1.10 RAPS PDU Transmission Mode in a Sub-ring

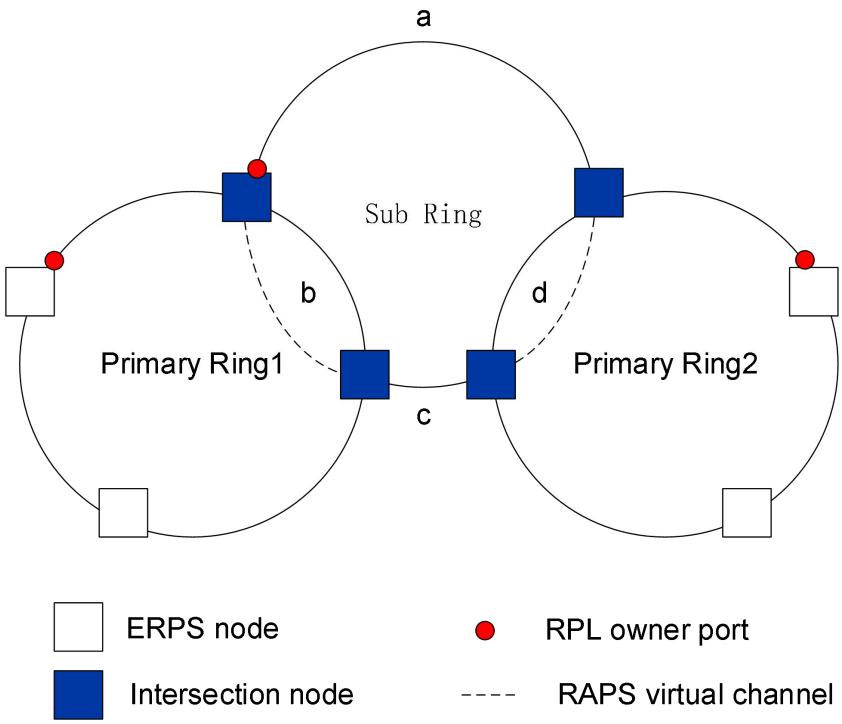
ERPSv2 supports single-ring and multi-ring topologies. In multi-ring topologies, both the virtual channel (VC) and non-virtual-channel (NVC) can be used to transmit RAPS PDUs in sub-rings.

- Virtual channel mode: RAPS protocol messages of sub-ring will run in the main ring through intersecting nodes. That is, the protocol message of the intersecting node does not terminate the sub-ring. The RPL owner port of the sub-ring blocks both RAPS PDUs and data traffic.
- NVC: RAPS PDUs in sub-rings are terminated on the interconnected nodes. The RPL owner port blocks data traffic but not RAPS PDUs in each sub-ring.

On the network shown in the Figure, a major ring is interconnected with two sub-rings. The sub-ring on the left has a VC, whereas the sub-ring on the right has an NVC.



On the network shown in the following Figure, links b and d belong to major rings 1 and 2 respectively; links a and c belong to the sub-ring. As links a and c are discontiguous, they cannot detect the status change between each other, so VCs must be used for RAPS PDU transmission.



The advantages and disadvantages of RAPS PDU transmission modes in sub-rings with VCs or NVCs are as follows:

- Virtual Channel
  - Advantages: It can be applied to special networking as shown above.
  - Disadvantage: The RAPS channel of a subnet is affected by the connected network topology, it requires VC resource reservation and controls VLAN assignment from adjacent rings.
- NVC
  - Advantage: Does not need to reserve resources or control VLAN assignment from adjacent rings.
  - Disadvantages: It cannot be applied to the special networking as shown above.

The RAPS message transmission mechanism on the ring network is as follows:

- When the port is in forwarding state, it normally forwards and accepts data messages and protocol messages;
- When the port is in a blocking state, it cannot forward and accept data messages and protocol messages, but it can accept and process protocol messages.
- When the port is in down state, it cannot forward and receive data messages and protocol messages normally.

7.2.1.11 Ring Level

The higher the ring network level, the greater the value. When the R-APS message needs to be transmitted across the ring, it can only be crossed by the ring with high rank to low rank.

7.2.1.12 ERPS Protocol State

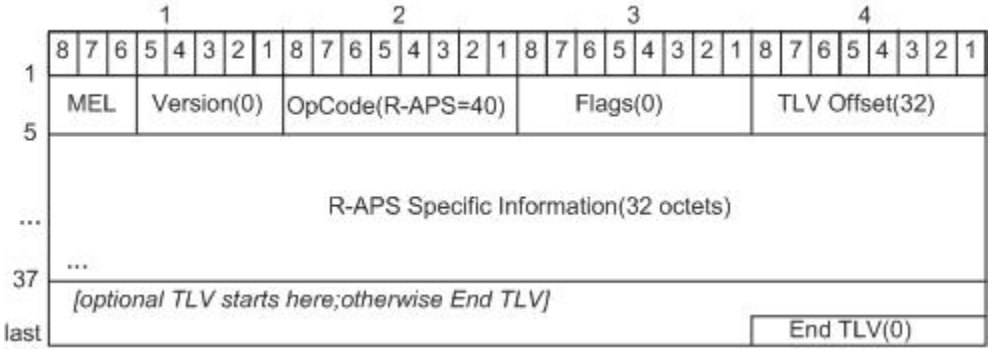
ERPS defines the following six states:

- Init status: ERPS ring is not enabled and is in Init status.
- Idle state: the ring enters a stable state after initialization. When the Owner node enters an Idle state, other nodes enter an Idle state. Among them, the RPL ports of Owner node and Neighbor node are blocked, that is, PRL is blocked; Owner node Schedule send (NR, RB) message.
- Protection state: when a link in a ring network fails, the ring is finally stabilized after protection switching. RPL ports of Owner node and Neighbor node are released, that is, PRL is released, which ensures that the whole ring network is still open. When a node in the link enters the Protection state, other nodes enter the Protection state.
- MS status: Enter MS status when manually switching traffic forwarding path. When a node in the link is subjected to MS operation, other nodes enter MS state.
- FS status: Enter FS status when forcibly switching traffic forwarding paths. When FS operation is performed on a node in the link, other nodes enter FS state.
- Pending state: Pending state is an unstable state, which is a transitional state when each state jumps.

When the loop is normal, it is in Idle state; The link is in Proctection state after failure.

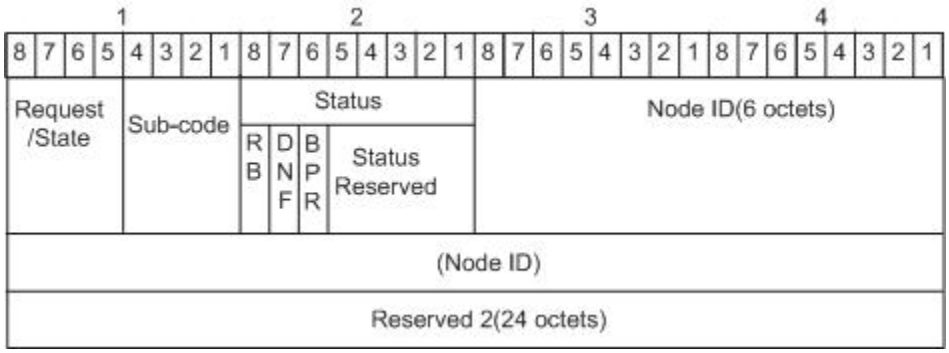
7.2.2 RAPS PDUs

There is only one message of ERPS protocol, namely RAPS PDU message. The basic format of RAPS PDU message is as follows:



- MEL: Identifying the level of R-APS messages, which is mainly applied to the determination of whether R-APS messages pass or not when they are transmitted across the ring. When the R-APS message needs to be transmitted across the ring, it can only be crossed by the ring with high rank to low rank.
- Version：ERPS protocol version, 0x00 identifies v1 version, 0x00 identifies v2 version;
- OpCode: fixed value 0x28, which identifies the R-APS message type;
- Flags: fixed value 0x00; this field is not used temporarily;
- TLV Offset: a fixed value of 0x20, which means that the TLV in the message starts after being offset by 32 bytes from this field;
- R-APS specific information: this field carries RAPS ring information and is the core field of RAPS PDU;
- End TLV: describes the information to be loaded in the message, with a fixed value of 0x00；

ERPSv2 version R-APS specific information message features are as follows



- Request/state(4 bits): the filling content of this field is indicated by the following table.

Field	Value	Description
Request/state	1101	Forced switch
	1110	Event
	1011	Signal fail(SF)

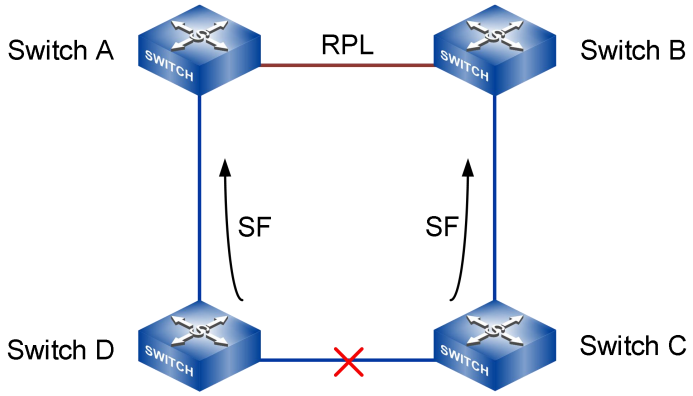
	0111	Manual switch(MS)
	0000	No request(NR)
	Other	Reserved for future international standardization

- Sub-code: when the fill value of Request/State is 1110, this field is filled with 0000 to indicate a refresh event; When the fill value of Request/State is other value, this field is filled with 0000 and the receiving end ignores this field.
- Status: status information;
  - The value of RB(RPL Blocked) is 1, indicating that the RPL port is in blocking state, and the value of RB (RPL blocked) is 0, indicating that the RPL port is in forwarding state.
  - The value of DNF(Do Not Flush) is 1, which means that FDB does not need to be refreshed, and the value of DNF (Do Not Flush) is 0, which means that FDB needs to be refreshed.
  - When the value of BPR(Blocked Port Reference) is 0, it means blocking the first port, and when it is 1, it means blocking the second port. This is mainly to unify the blocking rules of all devices on the whole ring network, so as to avoid the failure of ERPS function initialization and that it can not make all devices on the ring network enter the IDLE state correctly.
- Node ID: device node identification information, which is represented by the MAC address of the device;
- Reserved: This field will be filled with 0, which belongs to reserved field and will not be used for the time being.

7.2.3 ERPS Operation Mechanism

ERPS adopts the continuity detection defined in ITU-T G.8032/Y.1344 for link bidirectional forwarding detection, which can locate the fault point and detect whether the fault is unidirectional or bidirectional. ERPS judges the link status through the announced message, and makes corresponding processing. The control message types of ERPS mainly include SF(Signal Fail) and NR(No Request). If it is detected that the link fails to send and receive signals, the SF message will be sent. Send NR message when link recovery is detected. If the link state change is detected, three messages are sent continuously, and the subsequent messages are sent every five seconds.

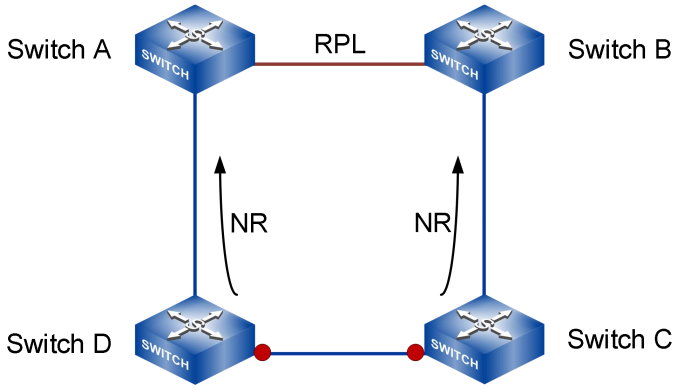
7.2.3.1 Link fault alarm mechanism



When a node in the link finds that any port belonging to ERPS ring is down, it will block the failed port and immediately send SF message to inform other nodes on the link that the failure has occurred. After receiving this message, other nodes will release the non-failed blocked port and refresh the MAC address table entry.

As shown in the above figure, when the link between Switch C and Switch D fails, Switch C and Switch D detect the link failure, block the failed port, and send SF messages periodically. After receiving SF messages, Switch A and Switch B release the previously blocked RPL port, switch the service to PRL, and complete the protection switching of the whole ring.

7.2.3.2 Link recovery mechanism



After the failed link is restored, block the port in the failed state before, start the Guard timer and send NR message to inform the Owner that the failed link has been restored. Owner node starts WTR timer after receiving NR message. if SF message is not received before the timer expires, after the timer expires, Owner node blocks RPL port and periodically sends (NR, RB) message outward. The fault recovery node releases the temporarily blocked fault recovery port after receiving the (NR, RB) message; The

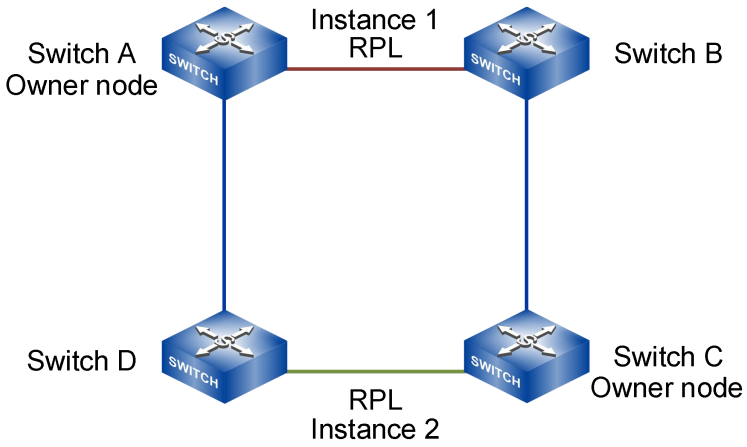
Neighbor node blocks the RPL port after receiving the (NR, RB) message, and the link is restored.

As shown in the above figure, when Switch C and Switch D detect that the link between them is restored, they temporarily block the ports that were in a failed state before and send NR messages. Upon receiving NR message, Switch A(Owner node) starts WTR timer. after the timer expires, it blocks RPL port and sends out (NR, RB) message. After receiving the (NR, RB) message, Switch C and Switch D release the temporarily blocked fault recovery port; Switch B(Neighbor node) blocks the RPL port after receiving the (NR, RB) message. The link is restored to the state before the failure.

There are two ways for the Owner node to recover the link.

- Revertive behaviour：Owner node will start the WTR/WTB timer after receiving NR message after fault elimination. Before the timer expires, if the Owner node does not receive the SF message, it switches the port status, blocks the RPL port, clears the MAC address table entry, and sends (NR, RB) messages. Other nodes release the non-fault blocked ports and clear their respective MAC address table entries. After the timer expires, it will switch back to the Idle state.
- Non-revertive behavior: after receiving NR message, the owner does not perform any action and keeps the previously set port status.

7.2.3.3 Multi-instance and load sharing mechanism



In the same ring network, data traffic of multiple VLANs may exist at the same time, and ERPS can realize load sharing of traffic, that is, traffic of different VLANs is forwarded along different paths.

ERPS ring network can be divided into control VLAN and data VLAN:

- Control VLAN: used to transmit ERPS protocol messages. Every ERPS instance has its own control VLAN.
- Data VLAN: Compared with control VLAN, data VLAN is used to transmit data

messages. Each ERPS instance has its own data VLAN, which is realized by configuring spanning tree instance.

By configuring multiple ERPS instances on the same ring network, different ERPS instances send traffic of different VLANs, so that the topology of data traffic of different VLANs in the ring network is different, thus achieving load sharing.

As shown in the above figure, Instance 1 and Instance 2 are two instances configured in a ring of ERPS, and the PRL of the two instances is different. The link between Switch A and Switch B is RPL of Instance 1, and Switch A is the Owner node of Instance 1. The link between Switch C and Switch D is RPL of Instance 2, and Switch C is the Owner node of Instance 2. Through configuration, different instances RPL block different VLAN, thus realizing load sharing of single ring.

7.2.3.4 Manual Configuration Mechanism

ERPS supports two levels of manual configuration, namely MS and FS.

- The MS allows the user to select ERPS ring member ports configured with MS mode in the current ring instance as blocking ports. The user configures ERPS < name > command manual-switch < east-port | west-port > command on the node where the port is located, and the node will send MS message to the outside. After receiving the MS message, other nodes will actively release ERPS ring member ports on their respective nodes, and finally stabilize the state that only ports configured by MS are blocked on the whole link. It should be noted that MS status can respond to link events, allowing switching to the corresponding status according to link events: in MS status, if other links fail, the node where the failure occurs will send SF messages outward, and other nodes will release ERPS ring member ports on their respective nodes after receiving SF messages, including MS configured blocking ports. At this time, the link can be switched to the protection state normally.
- The function of FS is similar to that of MS, except that in FS state, each node will not respond to link failure events, and the FS state will always remain unchanged.

7.3 Configure ERPS

7.3.1 Timer Configuration

7.3.1.1 Enter Timer Instance Configuration View

【Command】

erps timer config

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

erps timer config: this command is used to enter the ERPS timer configuration view.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps timer config
```

7.3.1.2 Create Timer Instance Name

【Command】

timer creat timer-name <*NAME*>

no timer creat timer-name <*NAME*>

【View】

ERPS timer instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS timer instance name.

【Description】

timer creat timer-name: this command is used to create a timer instance and specify the timer instance name.

By default, no configuration.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps timer config
Switch(config-erps-time)#timer creat timer-name 1
```

7.3.1.3 WTB Timer

【Command】

timer <*NAME*> set wtb <*interval*>

【View】

ERPS timer instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS timer instance name.

<interval>: wtb timer value, ranging from 1-12min.

【Description】

timer set wtb： this command is used to set the timing cycle of the TIMER instance WTB timer.

By default, it is 5min.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps timer config
Switch(config-erps- time)#timer 1 set wtb 1
```

7.3.1.4 WTR Timer

【Command】

timer <*NAME*> set wtr <*interval*>

【View】

ERPS timer instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS timer instance name.

<interval>: wtr timer value, ranging from 1-12min.

【Description】

timer set wtr: this command is used to set the timing cycle of the TIMER instance WTR timer.

By default, it is 5min.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps timer config
Switch(config-erps- time)#timer 1 set wtr 1
```

7.3.1.5 Guard Timer

【Command】

timer <*NAME*> set guard <*interval*>

【View】

ERPS timer instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS timer instance name.

<interval>: guard timer value, ranging from 10-2000ms.

【Description】

timer set guard: this command is used to set the timing cycle of the TIMER instance guard timer.

By default, it is 10ms.

【Instance】

```
Switch> enable
Switch#configure terminal
```

```
Switch(config)#erps timer config
Switch(config-erps- time)#timer 1 set guard 1000
```

7.3.1.6 Hold timer

【Command】

timer <*NAME*> set hold <*interval*>

【View】

ERPS timer instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS timer instance name.

<interval> : hold timer value, the range is 0-10s.

【Description】

timer set hold: this command is used to set the timing cycle of the TIMER instance hold timer.

By default, it is 0.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps timer config
Switch(config-erps- time)#timer 1 set hold 1
```

7.3.1.7 Display Timer Instance Information

【Command】

show erps {timer <*NAME*>| timer-all}

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS timer instance name.

【Description】

show erps timer: this command is used to display the instance information of the ERPS timer.

【Instance】

```
Switch#show erps timer 1
------------TIMER INSTANCE INFORMATION START------------
Timer        Name:1
WTR    Timer Value:1 min
WTB    Timer Value:5 min
Guard  Timer Value:10 ms
Hold   Timer value:0 ms
Hello  Timer value:5 s
------------TIMER INSTANCE INFORMATION END-------------
```

7.3.2 ERPS Ring Configuration

7.3.2.1 Enter ERPS ring configuration view

【Command】

erps ring config

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

erps ring config: this command is used to enter the ERPS ring configuration view.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps ring config
```

7.3.2.2 Create ERPS Ring Name

【Command】

ring creat ring-name <*NAME*> ring-id <*ring-id*>

no ring creat ring-name <*NAME*>

【View】

ERPS ring configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >: ERPS ring name.

< ring-id > : ERPS ring ID, the range is1-255.

【Description】

erps creat ring-name: this command is used to create the RING network and specify the RING name and RING ID.

By default, no configuration.

RING ID will be the last byte of the MAC destination of the RAPS message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps ring config
Switch(config-erps-ring)#ring creat ring-name 1 ring-id 1
```

7.3.2.3 Configure ERPS ring interface

【Command】

ring <*NAME*> set east-ifname <*if-name*> west-ifname <*if-name*>

【View】

ERPS ring configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >: ERPS ring name.

<if-name>: port name.

【Description】

ring set east-ifname: this command is used to configure the ring port for the specified RING loop.

By default, no configuration.

Notice:

1. ERPS ring ports can be normal physical ports or static aggregation groups.

2. ERPS ring port cannot be opened at the same time with other layer 2 protocols, when ERPS guard instance is not 0, it can be opened at the same time with MSTP.

3. ERPS ring ports can't be the same ports.

4. ERPS ring ports must be trunk ports and allow the ring instance VLAN to pass.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps ring config
Switch(config-erps-instance)#ring creat ring-name 1
Switch(config-erps-instance)#ring 1 set east-ifname ge1
west-ifname ge2
```

7.3.2.4 Configure ERPS ring level

【Command】

ring <*NAME*> set ring-level <*level*>

【View】

ERPS ring configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >: ERPS ring name.

<level> : ring grade, the range is 1-7.

【Description】

erps set ring-level: this command is used to configure the RING level for the specified RING Loop.

By default, the RING level is 1.

【Instance】

```
Switch> enable
Switch#configure terminal
```

```
Switch(config)#erps ring config
Switch(config-erps-ring)#ring creat ring-name 1
Switch(config-erps-ring)#ring 1 set ring-level 2
```

7.3.2.5 Display ERPS Ring Network Information

【Command】

show erps {ring <*NAME*>| ring-all}

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

< NAME >: ERPS ring name.

【Description】

show erps ring: this command is used to display the ERPS ring instance information.

【Instance】

```
Switch#show erps ring 1
--------------RING INSTANCE INFORMATION START---------------
Ring Name:1
East Port:ge2     Port Role:OTHER-PORT          Port State:BLOCK
West Port:ge1     Port Role:RPL-NEIGHBOR-PORT     Port State:BLOCK
Ring   ID:1      Ring Level:1               Ring  Role:Major Ring
---------------RING INSTANCE INFORMATION END-----------------
```

7.3.3 ERPS Instance Configuration

7.3.3.1 Enter ERPS Instance Configuration View

【Command】

erps instance config

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

erps instance config: this command is used to enter the ERPS instance configuration view.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
```

7.3.3.2 Create ERPS Instance Name

【Command】

erps creat erps-name <*NAME*>
no erps creat erps-name <*NAME*>

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

【Description】

erps creat erps-name: the command is used to create an ERPS instance and specify the instance name.

By default, no configuration.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
```

7.3.3.3 ConfigureERPS Instance ID

【Command】

erps <*NAME*> set instanceID <*instance*>
no erps <*NAME*> set instanceID

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

< instance > : MSTP instance number, the range is 0-16.

【Description】

erps set instanceID: the command is used to configure the ERPS protection instance (configured by the spanning tree).

By default, no configuration.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 set instanceID 1
```

7.3.3.4 Specify the Ring Instance Corresponding to the ERPS Instance

【Command】

erps <NAME> set ring <NAME>

no erps <NAME> set ring

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

< NAME > : ring instance name of ERPS

【Description】

erps set ring: this command is used to specify the ring instance corresponding to the ERPS instance (the ring instance is created and configured by the command in ring mode).

By default, no configuration.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 set ring 1
```

7.3.3.5 Specify the Timer Instance Corresponding to the ERPS Instance

【Command】

erps <*NAME*> set timer <*NAME*>

no erps <*NAME*> set timer

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

< NAME > : timer instance name of ERPS.

【Description】

erps set ring: this command is used to specify the timer instance corresponding to the ERPS instance (the timer instance is created and configured by the command in ring mode).

By default, no configuration.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 set timer 1
```

7.3.3.6 ERPS Instance Device Role

【Command】

erps <*NAME*> set role {rpl-owner | neighbor | interconnection} {east-ifindex | west-ifindex}
erps <*NAME*> set role other

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

【Description】

erps set role: this command is used to specify the role of the ERPS instance in the ring network.

By default, it is other.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 set role other
```

7.3.3.7 ERPS Instance Ring Role

【Command】

erps <*NAME*> set ring-role { major-ring| sub-ring}

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

【Description】

erps set ring-role: this command is used to specify the role of the ERPS instance in the ring network.

By default, it is a major-ring role.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 set role-ring major-ring
```

7.3.3.8 Major Instance Name of ERPS Instance

【Command】

erps <*NAME*> set major-instance-name <*NAME*>

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS subinstance name.

< NAME >：ERPS major instance name.

【Description】

erps set majority-instance-name: this command is used to set the major instance of ERPS for the specified subinstance of ERPS, and is executed only if the specified instance of ERPS is a subring.

By default, no configuration.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 set role-ring sub-ring
Switch(config-erps-instance)#erps 1 set major-instance-name 2
```

7.3.3.9 ERPS Instance Protocol Message Control VLAN

【Command】

erps <*NAME*> set raps-channel <*vlan*>

no erps <*NAME*> set raps-channel

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

< vlan > : VLAN carried in RAPS protocol message, the range is 1-4094.

【Description】

erps set raps-channel: this command is used to set the RAPS protocol message channel for the ERPS instance.

By default, it is 0 and invalid VLAN.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 set raps-channel 10
```

7.3.3.10 ERPS InstanceVirtual Channel

【Command】

erps <*NAME*> set virtual-channel {enable|disable}

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

【Description】

erps set ring-role: this command supports virtual channel, which is supported by enable and not supported by disable (note: virtual channel is not supported at present).

By default, it is default.

【Instance】

None

7.3.3.11 ERPS Instance Reverse Mode

【Command】

erps <*NAME*> set revertive {enable|disable}

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

【Description】

erps set revertive: command is used to configure the work mode of ERPS instance, enable is revertive mode and disable is irreversible mode.

By default, it is enable, that is reversible mode.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 set revertive disable
```

7.3.3.12 ERPS Instance Force-switch or Manual-switch

【Command】

```
erps <NAME> command { force-switch| manual-switch} {east-port | west-port}
```

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

【Description】

erps command: this command is used to perform forced or manual switch commands of ERPS instance.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 command force-switch west-port
```

7.3.3.13 ERPS Instance Clear Command

【Command】

erps <*NAME*> command clear

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

【Description】

erps command clear: this command is used to configure the ERPS instance to perform the clear command.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
```

```
Switch(config-erps-instance)#erps 1 command clear
```

7.3.3.14 ERPS Instance Enablement

【Command】

erps <*NAME*> {start|stop}

【View】

ERPS instance configuration view

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name.

【Description】

erps start|stop: this command is used to start or stop an ERPS instance.

By default, the ERPS instance is in the stop state.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#erps instance config
Switch(config-erps-instance)#erps creat erps-name 1
Switch(config-erps-instance)#erps 1 start
```

7.3.3.15 Display ERPS Instance Information

【Command】

show {erps <*NAME*>| erps-all}

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

< NAME >：ERPS instance name

【Description】

show erps: this command is used to display the ERPS instance information.

【Instance】

```
Switch#show erps 1

---------------ERPS INSTANCE INFORMATION START-----------------
ERPS  Name:1                          ERPS Version:1
ERPS-STATE:ERPS_PROTECTION            Device  Role:NEIGHBOR
InstanceID:0                          Channel Mode:NON-VRITUAL
ERPS  revert  mode:REVERTIVE
Major InstanceName:NULL
R-APS Vlan Channel:10
Data  Vlan Channel:NULL
WTR    Timer State:Stop
WTB    Timer State:Stop
Guard  Timer State:Stop
Hold   Timer State:Stop
Hello  Timer State:Running
Instance Run State:Running
-----------------RING INSTANCE INFORMATION START---------------
Ring Name:1
East Port:ge2    Port Role:OTHER-PORT           Port State:BLOCK
West Port:ge1    Port Role:RPL-NEIGHBOR-PORT    Port State:BLOCK
Ring   ID:1      Ring Level:1                Ring Role:Major Ring
----------------RING INSTANCE INFORMATION END-------------------
----------------TIMER INSTANCE INFORMATION START----------------
Timer        Name:1
WTR    Timer Value:1 min
WTB    Timer Value:5 min
Guard  Timer Value:10 ms
Hold   Timer value:0 ms
Hello  Timer value:5 s
---------------TIMER INSTANCE INFORMATION END-----------------
----------------ERPS INSTANCE INFORMATION END-----------------
```

8 Loop Detection Configuration

8.1 Overview

Loop detection is a single-node loop detection technology that uses periodic loop detection messages to detect the existence of loops between interfaces, downlink networks or devices, and between two interfaces of devices.

When a loop occurs on a network, broadcast, multicast, and unknown unicast packets are circulated on the network. This wastes network resources and can result in network breakdowns. Quickly detecting loops on a Layer 2 network is crucial for users to minimize the impact of loops on a network. LBDT and LDT help users check network connections and configurations, and control the looped interface.

Loop- Detect is just such a detection technique. It periodically sends a detection message from the interface to check whether the message is returned to the device, and then determines whether there are loops between the interface, the device's underlying network or the device, or between the two interfaces of the device. After a loop is detected, the device sends a trap to the NMS and records a log, and takes a preconfigured action on the looped interface (the interface is shut down by default) to minimize impact of the loop on the device and entire network.

8.2 Principles

Loop Detection technology is to periodically send a special Detection message from the interface, and then detect whether the message is returned to the device, and then determine whether there are loops between the interface, the device's downlink network or the device and the double interface of the device:

- If detection packets are received by the same interface, a loopback occurs on the interface or a loop occurs on the downstream network or device connected to the

interface.

- If detection packets are received by another interface on the same device, a loop occurs on the device or network connected to the interface.

After discovering the loop, the device will send an alarm to the network management and record the log, and close the interface at the same time to reduce the impact of the loop on the device and even the network. After the interface is closed, do not participate in any calculation or forwarding completely to prevent network storms.

After a certain period of time, if the device does not receive the detection message sent by the interface, the loop is considered to have been eliminated and the controlled interface will automatically return to the normal state. This process is called controlled interface automatic recovery. After the loop elimination, the recovery port can also be manually configured.

Controlled interface for Loop Detection based on VLAN Detection:

- If detection of this VLAN is canceled, the interface is restored automatically.
- If GVRP is not enabled on the interface and the interface is removed from the VLAN manually, the interface is restored automatically.

Loop Detection occurs in the interface or network, which will have an impact on normal business. However, Loop Detection is only a single-node Loop Detection technology, and it does not have the function of Loop breaking at the network level. After a loop is detected, you are advised to eliminate the loop immediately.

8.3 Global Enablement Configuration

【Command】

```
loop-detect enable
no loop-detect enable
```

【View】

Global configuration mode

【Default Level】

2. Configuration level

【Parameter】

None

【Description】

`loop-detect enable`: enable loop detection function.

no loop-detect enable: this command is used to disable loop-detect function.

【Instance】

```
Switch#configure terminal
Switch(config)#loop-detect enable
```

8.4 Forces the Port Closed by the Protocol to Open

【Command】

loop-detect force up

【View】

Ethernet port configuration mode

【Default Level】

2. Configuration level

【Parameter】

None

【Description】

loop - Detect force up: Forces to open the port closed by the protocol (this command does not save to the disk).

【Instance】

```
Switch>enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#loop-detect force up
```

8.5 Configure Protect VLAN

【Command】

loop-detect protect vlan <1-4094>

【View】

Ethernet port configuration mode

【Default Level】

2. Configuration level

【Parameter】

<1-4094> : VLAN ID range 1-4094.

【Description】

Loop -detect protect vlan <1-4094> : specify protected VLAN, enable port check.

【Instance】

```
Switch>enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#loop-detect protect vlan 1
```

8.6 Configure the Port Recovery Time

【Command】

loop-detect resume time <300-600>

【View】

Ethernet port configuration mode

【Default Level】

2. Configuration level

【Parameter】

<300-600> : port recovery time, range 300-600 seconds.

【Description】

loop-detect resume time <300-600> : restore port time.

【Instance】

```
Switch>enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#loop-detect resume time 300
```

8.7 Configure the Probe Packet Interval

【Command】

loop-detect tx-interval time <10-300>

【View】

Ethernet port configuration mode

【Default Level】

2. Configuration level

【Parameter】

<10-300> : detection packet detection interval, range 10-300 seconds.

【Description】

loop -detect tx-interval time <10-300> : time interval for sending probe packet.

【Instance】

```
Switch>enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#loop-detect tx-interval time 20
```

8.8 Displays Loop Detection Information

【Command】

show loop-detect [interface IFNAME]

【View】

Priviledged user mode

【Default Level】

1. View level

【Parameter】

IFNAME: port name.

【Description】

show loop-detect: shows the loop detection information of the device.

show loop-detect interface: displays loop detection information for the specified port.

【Instance】

```
Switch#show loop-detect
Loop-detection protocol: enable
Interface ge1
  isProtected       : Yes
  ProtectVlan list   : 1
  isLoop           : Yes
  LoopVlan list     : 1
  Port state        : No shutdown
  TX interval       : 10(s)
  TX up            : 1(s)
Interface ge22
```

```
isProtected       : Yes
ProtectVlan list   : 1
isLoop           : No
Port state        : Down
TX interval       : 10(s)
TX up            : 1(s)
```

9 IGMP Configuration

9.1 Overview

The Internet Group Management Protocol (IGMP) is a part of the TCP/IP protocol suite used to manage IPv4 multicast group membership. IGMP sets up and maintains membership between receiver hosts and directly connected multicast routers by exchanging IGMP messages between them. IGMP messages are encapsulated in IP packets.

IP multicast routing transmits packets from a source to a group of receivers. In the multicast communications model, the sender only needs to send data to a specified destination address and does not need to know the exact locations of the receivers. To forward multicast data packets to the receivers, the multicast router connected to the network segment of receiver hosts must know which receiver hosts are present on the network segment and ensure that these hosts have joined the specific group. IGMP implements this by setting up and maintaining membership between receiver hosts and directly connected multicast routers.

9.2 Principles

Currently, there are three versions of IGMP:

- IGMPv1
- IGMPv2
- IGMPv3

IGMPv1 defined the basic group member query and reporting process, IGMPv2 added the mechanism of queryer-electing and group member leaving on this basis, and the main function added in IGMPv3 is that members can specify to receive or not receive messages from certain multicast sources. The three versions are backward

compatible with each other in the process of evolution, so routers running the higher version of IGMP can identify the lower version of IGMP packets, although the format of each version of the protocol packet is different.

All IGMP versions support the any-source multicast (ASM) model. IGMPv3 can be directly applied to the source-specific multicast (SSM) model. IGMPv1 and IGMPv2, however, can be applied to the SSM model only when IGMP SSM mapping is configured.

The ASM model provides multicast distribution for group addresses only. A multicast group address acts as a collection of network services, and data published by any source to that group address gets the same service. After joining a multicast group, the receiver host can receive data sent to that group from any source.

The SSM model provides services for bound data streams for specific sources and groups, and when the receiver joins a multicast group, it can specify which sources only data will be received. After joining a multicast group, a host receives only the data sent to that group from the specified source.

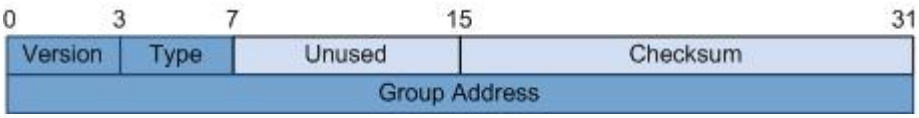
9.2.1 IGMPv1 Fundamentals

9.2.1.1 IGMPv1 Messages

IGMPv1 defines two types of message:

- General Query: A querier sends General Query messages to all hosts and routers on the local network segment to discover which multicast groups have members on the network segment.
- Report: Hosts send Report messages to multicast routers to request to join a multicast group or respond to General Query messages.

The format of IGMPv1 message is shown in the figure below.



- Version: IGMP version, value is 1.
- Type: Message Type. This field has the following two values:
  - 0x1: General Query message
  - 0x2: Report message
- Unused: in IGMPv1, the data line will be set to 0 during sending and will be ignored during receiving.
- Checksum: checksum of IGMP messages. It is the one's complement of the one's complement sum of the whole IGMP message (the entire IP payload). When

computing the checksum, a device initially sets the checksum field to 0. The sender computes the checksum and inserts it into this field. The receiver verifies the checksum before processing the message.

- Group Address. In a Membership Report message, this field is set to the address of the group that the member requests to join.

9.2.1.2 How IGMPv1 Works

IGMPv1 uses a query-report mechanism to manage multicast groups. When there are multiple multicast routers on a network segment, all multicast routers can receive Membership Report messages from hosts. Therefore, only one multicast router needs to send Query messages to the network segment. In IGMPv1, the only Multicast information retransmit was selected by Protocol Independent Multicast routing Protocol PIM as the IGMPv1 query, responsible for the group membership query of the segment.

The working mechanism of IGMPv1 can be divided into three aspects: general group query and response mechanism, new group member joining mechanism and group member leaving mechanism.

- Universal group query and response mechanism: By universal group query and response, IGMP query can know which multicast group has members within the network segment.
  - IGMP queryers periodically send generic group query packets with destination address 224.0.0.1 (representing all hosts and routers within the same network segment).
  - The group member that receives the query packet starts the timer, and the group member whose first timer time out sends a report packet to that group.
  - After the IGMP query receives the report message, the multicast forwarding delivery item is generated by the multicast routing protocol.
- New group member joining mechanism: multicast member actively sends report message for multicast to declare joining. After the IGMP query receives the report message, the multicast forwarding item is generated.
- Group member leaving mechanism: IGMPv1 has no message specifically defined for leaving a group. After a host leaves a group, it no longer responds to General Query messages.

9.2.2 IGMPv2 Fundamentals

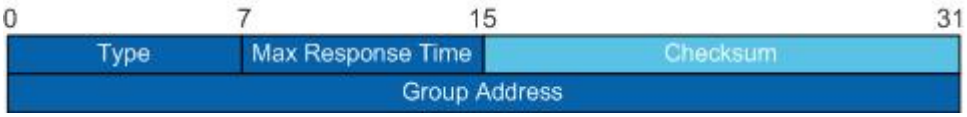
IGMPv2 works similarly to IGMPv1. The most significant difference between the two versions lies in the leave mechanism. When an IGMPv2 host leaves a group, it sends a Leave message to the IGMP querier. When the IGMP querier receives the Leave message, it sends a Group-Specific Query message to check whether the group has other members. If the IGMP querier does not receive any Report messages for the group within a specified period, it no longer maintains membership of the group. IGMPv2 enables an IGMP querier to know the groups that have no members on a local network segment and update group membership quickly. This leave mechanism reduces redundant multicast traffic on the network.

9.2.2.1 IGMPv2 Messages

IGMPv2 differs from IGMPv1 in the following ways:

- In addition to General Query and Report messages IGMPv2 defines two new message types:
  - Leave message: sent by a host to notify the querier on the local network segment that it has left a group.
  - Group-Specific Query message: sent by a querier to a specified group on the local network segment to check whether the group has members.
- IGMPv2 adds a new field to General Query messages: Max Response Time. This field can be configured to control hosts' response speed to Query messages.

The format of IGMPv2 message is shown in the figure below.



- Type: Message Type. This field has the following four values:
  - 0x11: Query message. IGMPv2 Query messages include General Query and Group-Specific Query messages.
  - 0x12: IGMPv1 Report message
  - 0x16: IGMPv2 Report message
  - 0x17: Leave message
- Maximum Response Time. After receiving a General Query message, hosts must respond with a Report message within the maximum response time. This field is valid only in IGMP Query messages. Checksum: checksum of IGMP messages. It is the one's complement of the one's complement sum of the whole IGMP

message (the entire IP payload). When computing the checksum, a device initially sets the checksum field to 0. The sender computes the checksum and inserts it into this field. The receiver verifies the checksum before processing the message.

- Group Address.
  - In the generic group query packet, this field is set to 0.0.0.0.
  - In a Group-Specific Query message, this field is set to the address of the queried group.
  - In a Report or Leave message, this field is set to the address of the group that the host wants to join or leave.

9.2.2.2 How IGMPv2 Works

IGMPv2 introduces the querier election and leave mechanisms.

- Query election mechanism: IGMPv2 USES an independent query election mechanism. When there are multiple multicast routers on the Shared network segment, the router with the smallest IP address becomes the query.
  - Initially, all multicast routers running IGMPv2 sent universal group query packets to all hosts and multicast routers within the network segment.
  - When the multicast router receives the generic group query packet sent by the other party, it compares the source IP address of the packet with its own interface address. The router with the smallest IP address becomes the querier, and the other routers are considered non-queriers.
  - After that, the IGMP query will be sent to all hosts and other multicast routers within the network segment, and the non-query will no longer send the universal group query message.
  - A Timer (i.e., Other Querier Present Timer) is started on the non-querier. If RouterB receive a Query message from the querier before the timer expires, it resets the timer. Otherwise, it triggers a querier election.
- **Leave mechanism:**
  - Members send outgoing messages to all multicast routers in the local network segment (destination address: 224.0.0.2).
  - When the query receives an exit message, it sends a specific set of query messages. Send intervals and send times can be configured by command. In addition, the querier starts the group membership timer. The timer length is the product of the group-specific query interval and count.
  - If there are other members in the network segment, these members will immediately send a report message to the multicast group after receiving a

specific group of query messages sent by the queriser. The Quaker will continue to maintain the group membership after receiving the report packet for the multicast group.

- If there are no other members of the multicast group in the network segment, the query will not receive a report message for the multicast group. After Timer-Membership times out, the query will delete the IGMP group table entries corresponding to the multicast group. When the multicast data with that multicast group reaches the query, the query will not forward it downstream.

9.2.3 IGMPv3 Fundamentals

IGMPv3 was developed to support the source-specific multicast (SSM) model. IGMPv3 messages can contain multicast source information so that hosts can receive data sent from a specific source to a specific group.

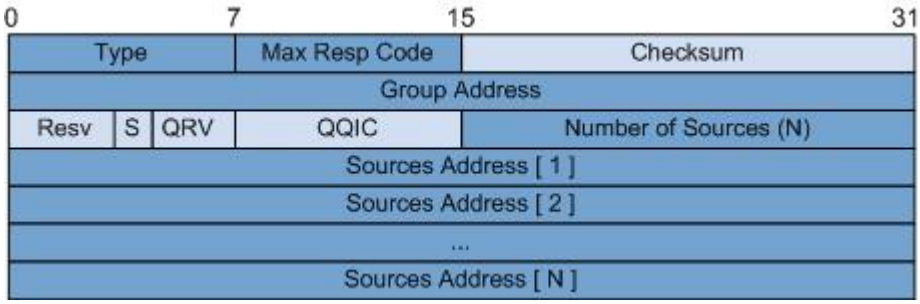
9.2.3.1 IGMPv3 Messages

IGMPv3 differs from IGMPv2 in the following ways:

- Unlike IGMPv2, IGMPv3 does not define a Leave message. Group members send Report messages of a specified type to notify multicast routers that they have left a group.
- In addition to General Query and Group-Specific Query messages, IGMPv3 defines another Query message type: Group-and-Source-Specific Query. A querier sends a Group-and-Source-Specific Query message to members of a specific group on a shared network segment, to check whether the group members want data from specific sources. A Group-and-Source-Specific Query message can carry one or more multicast source addresses.
- A Membership Report message contains the group that a host wants to join and the multicast sources from which the host wants to receive data. IGMPv3 supports source filtering and defines two filter modes: INCLUDE and EXCLUDE. Group-source mappings are represented as (G, INCLUDE, (S1, S2...)) or (G, EXCLUDE, (S1, S2...)). INCLUDE indicates that a host only wants to receive data sent from the listed multicast sources to group G. EXCLUD indicates that a host wants to receive data sent from all multicast sources except the listed ones to group G. When group-source mappings change, hosts add these changes to the Group Record fields in IGMPv3 Report messages and send IGMPv3 Report messages to the IGMP querier on the local network segment.
- An IGMPv3 Report message can carry multiple groups, whereas an IGMPv1 or IGMPv2 Report message can carry only one group. IGMPv3 greatly reduces the

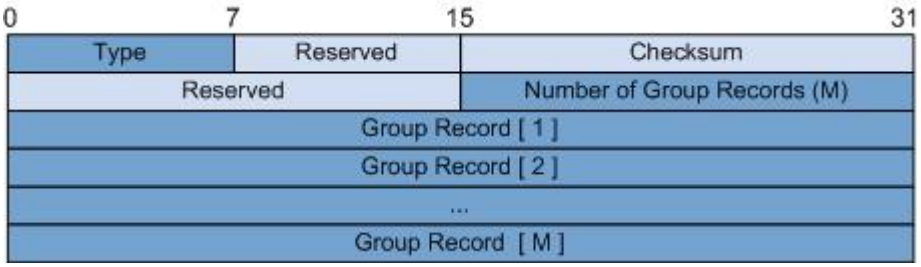
number of messages transmitted on a network.

The format of IGMPv3 query packets is shown in the figure below.



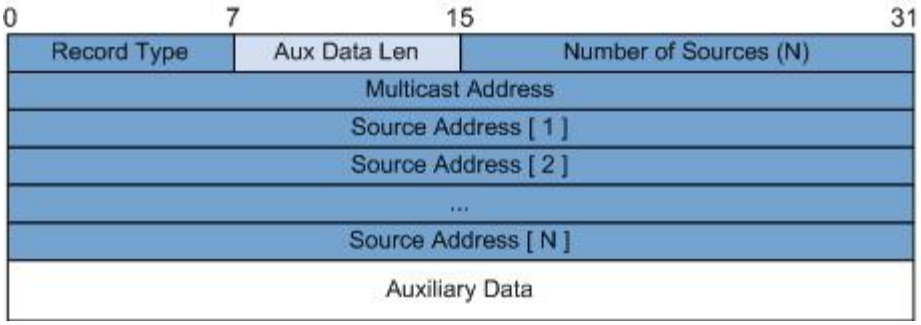
- Type: message type, value 0x11.
- Max Response Code: Maximum response time. After receiving a General Query message, hosts must respond with a Report message within the maximum response time.
- Checksum: checksum of IGMP messages. It is the one's complement of the one's complement sum of the whole IGMP message (the entire IP payload). When computing the checksum, a device initially sets the checksum field to 0. The sender computes the checksum and inserts it into this field. The receiver verifies the checksum before processing the message.
- Group Address. In a General Query message, this field is set to 0. In a Group-Specific Query or Group-and-Source-Specific Query message, this field is set to the address of the queried group.
- Resv: reserved field. This field is set to 0 by the sender and is ignored by the receiver.
- S: When the bit is 1, all other routers receiving the query packet do not start the timer refresh process, but the query packet does not inhibit the query selection process and the host side processing process of the router.
- QRV: If the field is not 0, it represents the query's Robustness Variable. This field is set to 0 if the robustness variable of the querier exceeds 7. When a router receives a Query message and finds that the QRV field is not 0, the router sets its robustness variable to the value of the QRV field. If the QRV field is 0, the router ignores this field.
- QQIC: IGMP query interval in seconds. When a non-querier receives a Query message and finds that the QQIC field is not 0, the non-querier sets its query interval to the value of the QQIC field. If the QQIC field is 0, the non-querier ignores this field.
- Number of Sources: the number of multicast Sources contained in a message. In a Group-and-Source-Specific Query message, this field is a non-zero integer. The number is limited by the MTU of the network over which the Query message is transmitted.
- Source Address: Multicast Source Address, the number of which is limited by the value size of the number of Sources field.

The format of IGMPv3 member report packets is shown in the figure below.



- Type: Message type, value 0x22.
- Reserved field. This field is set to 0 by the sender and is ignored by the receiver.
- Checksum: checksum of IGMP messages. It is the one's complement of the one's complement sum of the whole IGMP message (the entire IP payload). When computing the checksum, a device initially sets the checksum field to 0. The sender computes the checksum and inserts it into this field. The receiver verifies the checksum before processing the message.
- The number of group records contained in a message.
- Group record.

The format of the Group Record field is shown in the figure below.



- Record Type: type of group record. There are three main categories.
  - Current-State Record: sent by a host in response to a Query message to report the current state. This record type is one of the following two values:
  - 1. MODE_IS_INCLUDE: indicates that the host wants to receive multicast data sent from the listed source addresses to the specified multicast group address. The Report message is invalid if the source list is empty.
  - 2. MODE_IS_EXCLUDE: indicates that the host does not want to receive multicast data sent from the listed source addresses to the specified multicast group address.
  - Filter-Mode-Change Record: sent by a host when the filter mode changes. This record type is one of the following two values:
  - 1. CHANGE_TO_INCLUDE_MODE: indicates that the filter mode has changed from EXCLUDE to INCLUDE. The host wants to receive multicast

data sent from the new sources in the Source Address fields to the specified multicast group address. If the source list is empty, the host will leave the group.

- 2. CHANGE_TO_EXCLUDE_MODE: indicates that the filter mode has changed from INCLUDE to EXCLUDE. The host rejects multicast data sent from the new sources in the Source Address fields to the specified multicast group address.
- Source-List-Change Record: sent by a host when the source list changes. This record type is one of the following two values:
- 1. ALLOW_NEW_SOURCES: indicates that the Source Address fields in this group record contain additional sources from which the host wants to receive multicast data, sent to the specified multicast group address. If the filter mode is INCLUDE, the sources in the Source Address fields are added to the source list. If the filter mode is EXCLUDE, the sources in the Source Address fields are deleted from the source list.
- 2. BLOCK_OLD_SOURCES: indicates that the Source Address fields in this group record contain the sources from which the host no longer wants to receive multicast data. If the filter mode is INCLUDE, the sources in the Source Address fields are deleted from the source list. If the filter mode is EXCLUDE, the sources in the Source Address fields are added to the source list.

- Aux Data Len: auxiliary data length. IGMPv3 Report messages do not contain auxiliary data, so this field is set to 0.
- Number of Sources: The number of source addresses contained in this record.
- Multicast Address.
- Sources Address.
- Auxiliary Data. Reserved for subsequent extensions or versions of IGMP. There is no auxiliary data in the IGMPv3 report message.

9.2.3.2 How IGMPv3 Works

IGMPv3 differs from IGMPv2 and IGMPv3 in that it allows hosts to select specific multicast sources.

- **Joining a specific source and group**
  IGMPv3 Report messages have a destination address of 224.0.0.22, which represents all IGMPv3-capable multicast routers on the same network segment. A Report message contains Group Record fields, allowing hosts to specify the multicast sources from which they do or do not want to receive data when joining

a multicast group.

- Group-and-Source-Specific Query
  When receiving a report sent by a group member that changes the relationship between the multicast group and the source list, the IGMP query will send a specific source group query message. If members want to receive data from any source in the source list, they send Report messages. The IGMP querier updates the source list of the corresponding group according to the received Report messages.

9.2.4 IGMP SSM Mapping

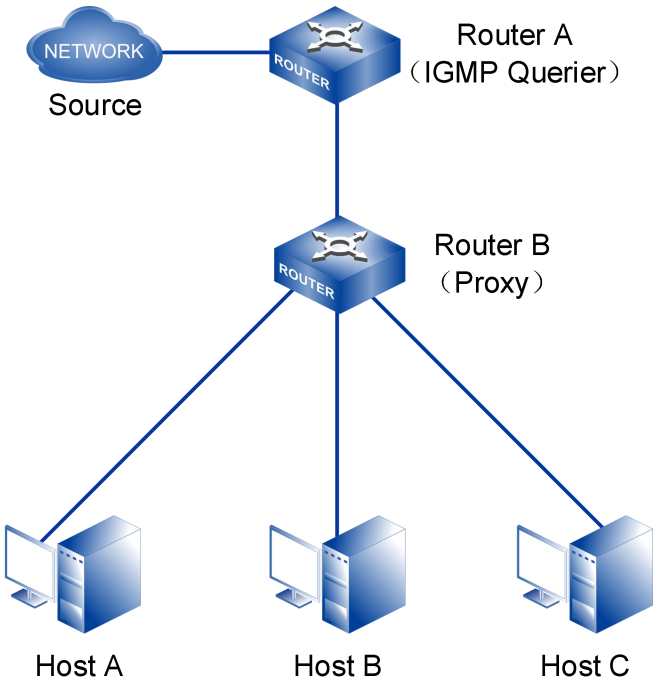
Source-specific multicast (SSM) requires multicast routers to know which multicast sources hosts specify when they join a multicast group. A host running IGMPv3 can specify multicast source addresses in IGMPv3 Report messages. However, hosts running IGMPv1 or IGMPv2 rely on the IGMP SSM mapping function to obtain the SSM service.

The mechanism of IGMP SSM Mapping is: by statically configuring SSM address Mapping rules on the router, information in IGMPv1 and IGMPv2 report packets is converted into corresponding information to provide SSM multicast service.

After the configuration of SSM Mapping rules, when the IGMP query receives the IGMPv1 or IGMPv2 report packets from the member host, it first checks the multicast group addresses carried in the paper, and then processes them separately according to the different inspection results.

- If the Multicast group is within the range of ANY-Source Multicast, then only ASM services are provided.
- If the multicast group is within the SSM group address range (the default is 232.0.0.0 ~ 232.255.255.255) :
  - If the router does not have the SSM Mapping rule corresponding to the multicast group, the SSM service cannot be provided and the article is discarded.
  - If there are SSM Mapping rules corresponding to the multicast group on the router, according to the rules, the information contained in the report packet (member, multicast group) will be mapped to (multicast group, INCLUDE, member) information, and SSM service will be provided.

9.2.5 IGMP Proxy



As shown in the figure above, in some simple tree network topologies, complex multicast routing protocols (such as PIM) do not need to be run on RouterB, a device connected to the user segment, and the transference of IGMP packets to the host leads to too many users being managed by RouterA. When a large number of receiver hosts exist on the network or many hosts frequently join and leave groups, a lot of IGMP Report/Leave messages are sent to RouterA, greatly increasing loads on RouterA.

By configuring the function of IGMP Proxy on RouterB, the above problems can be solved, the normal forwarding of group broadcast text can be achieved, and the processing pressure of RouterA can be alleviated at the same time.

The IGMP Proxy, also known as the IGMP Proxy, is typically deployed on a three-layer device between the access device (RouterA) and the member host, such as RouterB in the figure above. On one hand, the IGMP proxy device collects and processes IGMP Report/Leave messages from downstream hosts before forwarding the messages to the upstream access device. On the other hand, the IGMP proxy device substitutes the IGMP querier to send Query messages to downstream hosts to maintain group memberships, and forwards multicast data to hosts based on the group memberships. In this example, RouterB is a host for RouterA and an IGMP querier for the hosts.

IGMP proxy involves two types of interfaces:

- Upstream interface: an interface with IGMP proxy configured on the IGMP proxy device, which acts like a host. This interface is also called a host interface.
- Downstream interface: an interface with IGMP configured on the IGMP proxy device, which acts like a multicast router. This interface is also called a router interface.

9.2.5.1 How IGMP Proxy Works

An IGMP proxy device performs two types of behaviors: host behaviors and router behaviors.

Host behavior

When an upstream interface of the IGMP proxy device receives a Query message, the IGMP proxy device responds with a Report message according to the multicast forwarding table. When the multicast forwarding table changes, the upstream interface sends Report/Leave messages to the access device. An IGMP proxy device performs host behaviors in the following ways:

- When an upstream interface of the IGMP proxy device receives a Query message, the IGMP proxy device responds with a Report message according to the multicast forwarding table.
- When the IGMP proxy device receives a Report message for a group, it searches the multicast forwarding table for the group.
  - If the group is not found in the multicast forwarding table, the IGMP proxy device sends a Report message for the group to the access device and adds the group to the multicast forwarding table.
  - If the group is found in the multicast forwarding table, the IGMP proxy device does not send a Report message to the access device.
- When the IGMP proxy device receives a Leave message for a group, it sends a Group-Specific Query message through the downstream interface where the Leave message is received, to check whether this group has other members attached to the interface.
  - If there are no other members of this group attached to the interface, the IGMP proxy device deletes the interface from the forwarding entry of the group. The IGMP proxy device then checks whether the group has members on other interfaces. If so, the IGMP proxy device does not send a Leave message for this group to the access device. If not, the IGMP proxy device sends a Leave message for this group to the access device.
  - If the group has other members attached to the interface, the IGMP proxy device continues forwarding multicast data to the interface.

Router behavior

An IGMP proxy device generates multicast forwarding entries according to Report/Leave messages received on downstream interfaces, receives multicast data from the upstream access device, and forwards multicast data to downstream interfaces specified in the matching multicast forwarding entries. The mechanism of router behavior is consistent with that of IGMP.

9.2.5.2 Active/Standby and Active/Active Protection Mode

After enabling the IGMP proxy function on an upstream interface of an IGMP proxy device, you can configure an IGMP proxy interface protection mode to enhance link reliability. Two protection modes are supported:

- Main and standby mode: Configure IGMP Proxy backup on the other interface of the device, and protect the main IGMP Proxy interface. When the main IGMP Proxy interface fails, the upstream interface supporting the table items is switched to the standby IGMP Proxy interface to ensure rapid convergence of traffic. When the main IGMP Proxy interface fails to recover, the upstream interface of the table entries will cut back to the main IGMP Proxy interface.
- Master master mode: IGMP Proxy is configured on other interfaces on the device, and multiple main IGMP Proxy interfaces are load sharing according to the hash algorithm. IGMP Proxy devices send incoming/outgoing packets to each main IGMP Proxy interface, and all the main IGMP Proxy interfaces protect each other. If one of the main IGMP Proxy interfaces fails, the upstream interface of the table entries will switch to other main IGMP Proxy interfaces to ensure fast convergence of traffic. When the master IGMP Proxy interface fails to recover, the upstream interface of the table item is loaded from one master link to multiple master links according to the hash algorithm.

9.2.5.3 Automatic Switchback and Delayed Switchback

- Automatic switchback: Traffic is switched back immediately after the primary interface recovers.
- Delayed switchback: Traffic is not switched back immediately after the primary interface recovers. Instead, traffic switchback is performed after a delay, which prevents frequent traffic switchover and switchback in the case that the primary interface flaps. You can manually trigger traffic switchback before the delay expires.

9.3 IGMP Configuration

9.3.1 Configure IGMP Basic Functions

9.3.1.1 IGMP Interface Enable

【Command】

ip igmp

no ip igmp

【View】

VLAN-IF Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip igmp: This command is used to enable IGMP on the interface.

no ip igmp: this command is used to disable IGMP on the interface.

By default, IGMP on the interface is disabled.

Configuration of other IGMP features on an interface takes effect only if IGMP is enabled on that interface.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp
```

9.3.1.2 IGMP Versions

【Command】

ip igmp version VERSION-NUMBER

no ip igmp version

【View】

VLAN-IF Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

VERSION-NUMBER:: represents the version number of IGMP, and the value range is 1-3.

【Description】

ip igmp version: this command is used to configure the version of IGMP on the interface.

no ip igmp version: this command is used to restore the IGMP to the default value. By default, the version of IGMP is IGMPv3.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp version 2
```

9.3.1.3 Static Multicast

【Command】

ip igmp static-group <group-address> [source <source-address> | ssm-map]

no ip igmp static-group <group-address> [source <source-address> | ssm-map]

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

group-address: specify multicast group address with values ranging from 224.0.1.0 to 239.255.255.255.

source-address: specifies the address of the multicast source.

ssm-map: obtain the multicast source address through ssm - mapping function.

【Description】

ip igmp static- group: This command is used to configure the interface to statically join a multicast group or multicast source group.

no ip igmp static-group: this command is used to restore to the default value. By default, the interface does not statically join any multicast group or multicast source group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp static 225.1.1.1 source 192.168.1.10
```

9.3.1.4 Illegal Multicast Group

【Command】

ip igmp access-group <ACL-NUMBER | ACL-NAME>

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

acl - number: A standard ACL number with values ranging from 1 to 99.

acl-name：extend the ACL name.

【Description】

ip igmp access-group: This command is used to configure an illegal multicast group scope. The ACTION for the ACL must be deny. If the action is permit, the mismatched multicast group is also considered a valid multicast group.

no ip igmp access-group: this command is used to unrestrict the scope of an illegal multicast group.

By default, the range of multicast groups that an interface can learn is unrestricted.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 1300 deny 225.1.1.0 0.0.0.255
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp access-group 99
```

9.3.1.5 Multicast Groups do not Age

【Command】

```
ip igmp groups no-timeout
no ip igmp groups no-timeout
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip igmp groups no-timeout: this command is used to configure the multicast group not to age, and to leave normally when the leave message is received.

no ip igmp groups no-timeout: this command is used to cancel the multicast group without aging.

By default, multicast groups age normally.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)# ip igmp groups no-timeout
```

9.3.1.6 Display IGMP Multicast Information

【Command】

```
show ip igmp groups [<IFNAME> | <GROUP-ADDRESS> | detail ]
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: vlanif interface.

GROUP-ADDRESS：Ipv4 multicast addresses.

detail: outputs the details of a multicast group.

【Description】

View the running condition of the specified parameter or the entire multicast group.

【Instance】

```
Switch#show ip igmp groups detail
IGMP Connected Group Membership, Total is 1
Interface:     vlanif1
Group:       255.1.1.1
Uptime:      01:09:31
Group mode:   Exclude (Expires: 00:03:56)
Last reporter:   192.168.1.11
Source list is empty
```

9.3.1.7 Displays IGMP Interface Information

【Command】

show ip igmp inteface <IFNAME>

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: vlanif interface

【Description】

View the configuration and running condition of the interface with specified parameters or all IGMP enabled.

【Instance】

```
Switch#show ip igmp interface vlanif1
Interface vlanif1 (Index 3)
 IGMP Enabled, Active, Querier, Configured for version 2
 L3 mcast is not enabled on this interface
 Internet address is 192.168.1.254
 IGMP interface has 1 group-record states
 IGMP activity: 54 joins, 0 leaves
 IGMP query interval is 125 seconds
 IGMP Startup query interval is 31 seconds
 IGMP Startup query count is 2
 IGMP querier timeout is 255 seconds
```

```
IGMP max query response time is 10 seconds
Last member query response interval is 1000 milliseconds
Group Membership interval is 260 seconds
IGMP Last member query count is 2
L2 mcast is not enabled on this interface
IGMP Snooping is globally disabled
IGMP Snooping is not enabled on this interface
IGMP Snooping fast-leave is not enabled
IGMP Snooping querier is not enabled
IGMP Snooping report suppression is enabled
```

9.3.1.8 Displays IGMP Global Information

【Command】

show ip igmp instance

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show ip igmp instance: View device multicast group global information, such as multicast group limit, etc.

【Instance】

```
Switch#show ip igmp instance
IGMP global Information
 IGMP global limit states is unset
 IGMP global limit states count is currently 0
 IGMP-Snooping send source-address is 192.168.0.1
 Interface count is 28
```

9.3.2 Adjust IGMP Performance

9.3.2.1 IGMP Message with RA Option

Normally, when a network device receives a message, only the message whose destination IP address is the interface address of the device will be sent to the corresponding protocol module for processing. If the destination address of the protocol message is not the interface address of the device, such as the IGMP protocol message, because its destination address is the multicast address, in this case, it cannot be sent to the IGMP protocol module for processing, resulting in the failure to maintain the normal group membership. To solve such problems, the Router-Alert option was created. If an IP message header carries router-alert option, the device receives such a message and sends it directly to the appropriate protocol module for processing without checking the destination address.

For compatibility, the current switch will be sent to the IGMP protocol module by default after receiving an IGMP message, regardless of whether its IP header contains router-Alert option. In order to improve device performance and reduce unnecessary costs, and for protocol security, switches can also be configured to discard IGMP message that do not carry Router-Alert option.

【Command】

```
ip igmp ra-option
no ip igmp ra-option
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

`ip igmp ra-option`: this command is used to configure the interface to discard igmp message that do not carry the Router-Alert option.

`no ip igmp ra-option`: this command is used to restore the default value.

By default, devices do not check for the Router-Alert option, that is, they send all IGMP message they received to the upper protocol for processing, whether or not they carry the Router-Alert option.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp ra-option
```

9.3.2.2 The number of times an IGMP query is started

【Command】

```
ip igmp startup-query-count <2-10>
no ip igmp startup-query-count
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

<2-10> : specifies the number of times an IGMP query is started, with a value range of 2-10.

【Description】

ip igmp startup-query-count: this command is used to configure the number of start queries for the IGMP querier on the interface.

no ip igmp startup-query-count: this command is used to restore to the default value.

By default, the IGMP query is started 2 times.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp startup-query-count 5
```

9.3.2.3 The Start Query Interval for an IGMP Query

【Command】

```
ip igmp startup-query-interval <1-18000>
no ip igmp startup-query-interval
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

<1-18000> : specifies the start query interval for the IGMP query, in the range 1-18000 in seconds.

【Description】

ip igmp startup-query-interval: this command is used to configure the start query interval for the IGMP query on the interface.

no ip igmp startup-query-interval: this command is used to restore to the default value.

By default, the IGMP query starts at a quarter of the time it took to send the IGMP universal group query message. The time interval of sending IGMP universal group query message is 125 seconds, then the start query interval of IGMP querier is 125÷4=31 (seconds).

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp startup-query-interval 20
```

9.3.2.4 The Robustness Factor of the IGMP Query

【Command】

```
ip igmp robustness-variable <2-7>
no ip igmp robustness-variable
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

<2-7> : specifies the robustness factor of the IGMP query, with a value range of 2-7. This coefficient is used to specify the default value of the number of times an IGMP query message is sent by the IGMP query at startup, and the number of times an

IGMP query message is sent by the IGMP query after the IGMP query receives the message leaving the group.

【Description】

ip igmp robustness - variable: this command is used to configure the robustness coefficient of the IGMP query on the interface.

no ip igmp robustness-variable: this command is used to restore to the default value.

By default, the IGMP query has a robustness factor of 2.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp robustness-variable 3
```

9.3.2.5 IGMP universal group queries the time interval of packets

【Command】

ip igmp query-interval <1-1800>

no ip igmp query-interval

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

<1-1800> : specifies the time interval between sending IGMP universal group query messages, with a value range of 1-18000, in seconds.

【Description】

ip igmp query-interval: this command is used to configure the interval at which IGMP universal group query messages are sent on the interface.

no ip igmp query-interval: this command is used to restore the default.

By default, IGMP universal group query messages are sent at an interval of 125 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
```

```
Switch(config-vlanif1)#ip igmp query-interval 240
```

9.3.2.6 The IGMP universal group queries the maximum response time of message

【Command】

```
ip igmp query-max-response-time <1-240>
no ip igmp query-max-response-time
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

<1-240> : specifies the maximum response time of IGMP universal group query message, and the value range is 1-240, in seconds.

【Description】

ip igmp query-max-response-time: this command is used to configure the maximum response time for IGMP universal group queries on the interface.

No ip igmp query-max-response-time: this command is used to restore the default.

By default, the maximum response time of IGMP universal group query messages is 10 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp query-max-response-time 20
```

9.3.2.7 The Lifetime of Other IGMP Queries

【Command】

```
ip igmp querier-timeout <60-300>
no ip igmp querier-timeout
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

<60-300> : specifies the duration of the IGMP other queries, in the range of 60-300 in seconds.

【Description】

ip igmp querier-timeout: this command is used to configure the lifetime of IGMP other queries on the interface.

no ip igmp querier-timeout: this command is used to restore to the default value.

By default, the existence time of other IGMP queriers = the time interval of sending IGMP universal group query message × the robustness coefficient of IGMP queriers + the maximum response time of IGMP universal group query ÷2.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp querier-timeout 180
```

9.3.2.8 Number of IGMP query packets for a specific group

【Command】

```
ip igmp last-member-query-count <2-7>
no ip igmp last-member-query-count
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

<2-7> : specifies the number of query message to send IGMP specific groups, and the value range is 2-7.

【Description】

ip igmp last-member-query-count: this command is used to configure the number of sending query message of IGMP specific groups on the interface.

no ip igmp last-member-query-count: command is used to restore to the default value.

By default, the number of sending IGMP specific group query message is 2.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp last-member-query-count 3
```

9.3.2.9 The time interval for a particular group of IGMP query packets

【Command】

ip igmp last-member-query-interval <1000-25500>

no ip igmp last-member-query-interval

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

<1000-25500> : specifies the time interval for sending IGMP specific group of query messages. The value range is 1000-25500, in milliseconds.

【Description】

ip igmp last-member-query-interval: the command is used to configure the interval at which the IGMP specific group query message is sent on the interface.

no ip igmp last-member-query-interval: the command is used to restore the default.

By default, IGMP specific group query message are sent at an interval of 1000 ms.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp last-member-query-interval 2000
```

9.3.2.10 Fast Aging ACL Group

【Command】

ip igmp immediate-leave group-list <ACL-NUMBER | ACL-NAME>
no ip igmp immediate-leave

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

acl - number: a standard ACL number with values ranging from 1-99 or 1300-1999.

acl-name：extend ACL name.

【Description】

ip igmp immediate-leave group-list: This command is used to configure the address range of the fast-leaving ACL group, the ACL action must be permit.

no ip igmp immediate-leave: this command is used to restore to the default.

By default, when the interface works in Version 2 and Version 3, after receiving IGMP leave message, a specific group of query message will be sent first to determine whether the multicast member table entries are aging. With this feature configured, multicast member table entries can be aged immediately if the group address specified by the message is left within the group address range specified by the ACL.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 1300 permit 225.1.1.0 0.0.0.255
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp immediate-leave group-list 1300
```

9.3.2.11 IGMP Message Source Address and Receive Interface Subnet Restrictions

【Command】

ip igmp offlink
no ip igmp offlink

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip igmp offlink: this command is used to remove the restriction that the source address of an IGMP message must be in the same subnet as the receiving interface, except for querying message and leaving message.

no ip igmp offlink: this command is used to restore to the default value.

By default, the source address of an IGMP Snooping message must be on the same subnet as the receiving interface. Once the restriction is removed, the source address of the message will be considered valid as long as it passes RPF check.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp offlink
```

9.3.2.12 Do not participate in the IGMP query election

【Command】

ip igmp no-querier-election

no ip igmp no-querier-election

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip igmp no-querier-election: This command is used to configure the device not to participate in the IGMP query election.

no ip igmp no-querier-election: This command is used to configure the device not to participate in the IGMP query election.

By default, participate in the IGMP query election.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp no-querier-election
```

9.3.3 Configure IGMP Limit

9.3.3.1 Limited Number of Multicast Groups

【Command】

ip igmp limit VALUE [except <ACL-NUMBER | ACL-NAME >]

no ip igmp limit

【View】

Global configuration mode

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

VALUE: the maximum number of multicast groups allowed to be added by the global or interface, ranging from 1 to -1024.

ACL- NUMBER: standard ACL number, ranging from 1-99 or 1300-1999.

ACL-NAME：extend ACL name.

【Description】

ip igmp limit: this command is used to configure the maximum number of multicast groups that are allowed to be added to the global or interface.

no ip igmp limit: this command is used to restore to the default value. In the process of working, the device first determines whether it exceeds the global limit, then determines whether it exceeds the interface limit, or ignores the new multicast group learning.

An ACL of type permit can be referenced by the except parameter, indicating that there are no restrictions on the number of multicast groups within the range specified by the ACL.

By default, there is no limit to the number of multicast groups that can be added to a global or interface.

Notice

When the configured limit value is less than the number of established multicast groups on the global or current interface, the system will not automatically delete the additional multicast groups.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip igmp limit 1000
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp limit 100
```

9.3.4 Configure IGMP SSM Mapping

9.3.4.1 Enable Global IGMP SSM Mapping

【Command】

ip igmp ssm-map enable

no ip igmp ssm-map enable

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip igmp ssm-map enable: this command is used to enable the IGMP SSM Mapping function globally.

no ip igmp ssm-map enable: this command is used to disable the global IGMP SSM Mapping function.

By default, the IGMP SSM Mapping function is disabled.

IGMPv1/IGMPv2 cannot specify the multicast source in the report message, so IGMP SSM Mapping technology is required for compatibility. This feature provides SSM services for the interface that receives vesion1 and version2 igmp report packets, with the source address specified by the ip igmp ssm-map static command.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip igmp ssm-map enable
```

9.3.4.2 Configure IGMP SSM-Map Static Multicast

【Command】

ip igmp ssm-map static <acl-number | acl-name> <source-address>
no ip igmp ssm-map static <acl-number | acl-name> <source-address>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

acl - number: a standard ACL number with values ranging from 1～ 99 or 1300～ 1999.

acl-name：extend the ACL name.

source-address: the static mapping source address for SSM mapping.

【Description】

ip igmp ssm-map static: this command is used to configure the IGMP SSM Mapping rule.

no ip igmp ssm-map: this command is used to delete the IGMP SSM Mapping rule. IGMP SSM Mapping rules are not configured by default.

IGMPv1/IGMPv2 cannot specify the multicast source in the report message, so IGMP SSM Mapping technology is required for compatibility. This feature needs to work with ip igmp ssm-map enable.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 1300 deny 225.1.1.0 0.0.0.255
Switch(config)#ip igmp ssm-map enable
```

```
Switch(config)#ip igmp ssm-map static 1300 192.168.1.10
```

9.3.4.3 Display IGMP SSM-Map Mapping Rules

【Command】

```
show ip igmp ssm-map [A.B.C.D]
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

A.B.C.D: multicast group address.

【Description】

View the specified source multicast group or all IGMP SSM-Map mapping rule information.

【Instance】

```
Switch#show ip igmp ssm-map
```

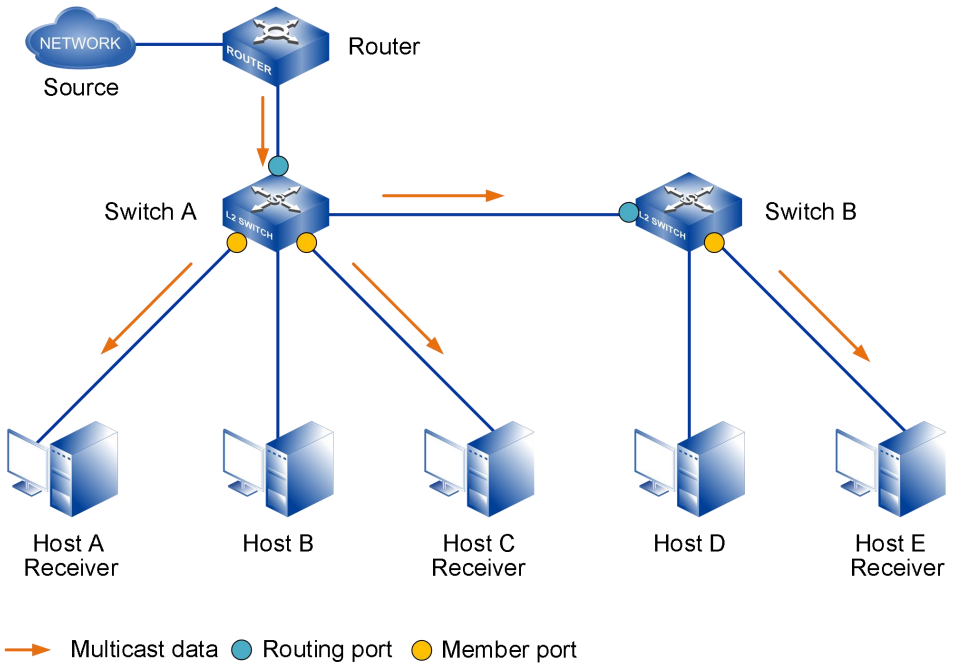
10 IGMP Snooping Configuration

10.1 Overview

Internet Group Management Protocol (IGMP) snooping is a Layer 2 IPv4 multicast protocol. It listens to multicast protocol packets exchanged between the upstream Layer 3 multicast device and downstream hosts to maintain outbound interfaces of multicast packets. Based on the outbound interface information, IGMP snooping manages and controls the multicast forwarding at the data link layer.

10.2 Principle Description

IGMP Snooping is the basic function of layer 2 multicast. It can forward and control the multicast data in the data link layer. When IGMP protocol messages transmitted between host and upstream layer 3 device pass through the layer 2 multicast device, IGMP Snooping analyzes the information carried by the message, establishes and maintains the layer 2 multicast forwarding based on the information, and then directs the multicast data to be forwarded as needed at the data link layer.



As shown in the figure above, when the multicast data is forwarded from the Router of the layer 3 multicast device, Switch, a layer 2 multicast device at the access edge, is responsible for forwarding the multicast data to the user's host, so that the user can watch the programs on demand. When Switch does not run IGMP Snooping, the multicast data is broadcast at layer 2. When Switch runs IGMP Snooping, the multicast data is not broadcast at layer 2, but is instead sent by Switch to the specified receiver.

With IGMP snooping configured, the switch listens to IGMP messages exchanged between the router and hosts. It analyzes packet information (such as packet type, group address, and receiving interface) to set up and maintain a Layer 2 multicast forwarding table, based on which it forwards multicast packets.

10.2.1 IGMP Snooping Relative Ports

- Router port

  A router port receives multicast packets from a Layer 3 multicast device such as a designated router (DR) or IGMP querier.

  - The router port generated by the protocol is called the dynamic router port. An IGMP universal group query message with a source address of 0.0.0.0 or an interface to a PIM Hello message(a message sent by the PIM interface of a layer 3 multicast device to discover and maintain neighborhood relationships) is considered a dynamic router port.
  - A static router port is manually configured.

- Member port
  A member port is a user-side port connected to group members. A Layer 2 multicast device sends multicast data to receiver hosts through member ports.
  - The member ports generated by the protocol are called dynamic member ports. The interface to which IGMP Report message are received will be identified by the layer 2 multicast device as a dynamic member port.
  - A static member port is manually configured.

The router port and the member port are the important information in the layer 2 multicast forwarding item: the egress interface. The router port is equivalent to the upstream interface, and the member port is equivalent to the downstream interface. The port learned through the protocol message corresponds to the dynamic table item; Manually configured ports correspond to static table entries.

In addition to the outbound interfaces, each entry includes multicast group addresses and VLAN IDs.

- Multicast group addresses can be multicast IP addresses or multicast MAC addresses mapped to multicast IP addresses. Following the IP address forwarding mode can avoid the problem of address duplication in MAC address forwarding mode.
- VLAN number that specifies the layer 2 broadcast domain range. If a multicast VLAN function is used, the ingress VLAN number is the number of the multicast VLAN, and the egress VLAN number is the user VLAN number of the host. Otherwise the ingress VLAN number and the egress VLAN number are the VLAN Numbers of the host.

10.2.2 Implementation

A Layer 2 multicast device running IGMP snooping processes received IGMP protocol packets in different ways to set up Layer 2 multicast forwarding entries.

Processing of different messages by IGMP Snooping:

- Universal Group Query (IGMP Universal Group Query Message)
  The IGMP querier periodically sends General Query messages (with destination address 224.0.0.1) to all hosts and routers on the local network segment, to check which multicast groups have members on the network segment. When the interface receives the IGMP general group query message, it will forward to all other interfaces in the VLAN except the receiving interface, and the receiving interface will be processed as follows:
  - If the port is not in the router port list, the Layer 2 multicast device adds it to the list and starts the aging timer.

- If the port is included in the router port list, the Layer 2 multicast device resets the aging timer of the router port.

- Member report (IGMP report message)

  There are two cases: after receiving the IGMP universal group query message, members respond to the IGMP report message; Members actively send IGMP report message to the IGMP query to declare membership in the multicast group. When the interface receives the IGMP report message, it forwards to all router ports in the VLAN. The multicast group address to be added by the host is resolved from the message, and the receiving interface is processed as follows:

  - If the multicast group matches no forwarding entry, the Layer 2 multicast device creates a forwarding entry, adds the port to the outbound interface list as a dynamic member port, and starts the aging timer.
  - If the multicast group matches a forwarding entry but the port is not in the outbound interface list, the Layer 2 multicast device adds the port to the outbound interface list as a dynamic member port, and starts the aging timer.
  - If the multicast group matches a forwarding entry and the port is in the router port list, the Layer 2 multicast device resets the aging timer.

- Member leaving multicast group (IGMP leaving message, IGMP specific group query message /IGMP specific source group query message)

  There are two phases: members running IGMPv2 or IGMPv3 send IGMP exit messages to inform the IGMP query that they have left a multicast group; After the IGMP query receives the IGMP leave message, the address of the multicast group is parsed from it, and the IGMP specific group query message /IGMP specific source group query message is sent to the multicast group through the receiving interface. The treatment is as follows:

  The Layer 2 multicast device determines whether the multicast group matches a forwarding entry and whether the port that receives the message is in the outbound interface list.

  - If no forwarding entry matches the multicast group or the outbound interface list of the matching entry does not contain the receiving port, the Layer 2 multicast device drops the IGMP Leave message.
  - If the multicast group matches a forwarding entry and the port is in the outbound interface list, the Layer 2 multicast device forwards the IGMP Leave message to all router ports in the VLAN.

  The following assumes that the port receiving an IGMP Leave message is a dynamic member port. Before the aging time of the member port expires:

- If the port receives IGMP Report messages in response to the IGMP Group-Specific/Group-Source-Specific Query message, the Layer 2 multicast device knows that the group has members connected to the port and resets the aging timer.
- If the port receives no IGMP Report message in response to the IGMP Group-Specific/Group-Source-Specific Query message, no member of the group exists under the port. When the aging time is reached, the Layer 2 multicast device deletes the port from the outbound interface list.

Upon receiving a PIM Hello message, a Layer 2 multicast device forwards the message to all ports except the one that receives the Hello message. The Layer 2 multicast device processes the receiving port as follows:

- If the port is included in the router port list, the Layer 2 multicast device resets the aging timer of the router port.
- If the port is not in the router port list, the Layer 2 multicast device adds it to the list and starts the aging timer.

If a static router port is configured, the layer 2 multicast device will also forward IGMP reports and leave message to the static router port. If a static member port is configured, that interface is added as an egress interface in the forward entry.

When the layer 2 multicast forwarding item is established on the layer 2 multicast device, when the layer 2 multicast device receives the multicast data message, it will look for the corresponding "egress interface information" of the forwarding item according to the VLAN to which the message belongs and the destination address of the message (namely, the multicast group address). If there is, the message is sent to the member port of the multicast group and the router port;

If it does not exist, the message is discarded or broadcast within the VLAN.

10.3 Configure IGMP Snooping

10.3.1 IGMP Snooping Enablement

【Command】

```
ip igmp snooping
no ip igmp snooping
```

【View】

Global configuration mode

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip igmp snooping: this command is used to enable IGMP snooping on global or VLAN interfaces.

no ip igmp snooping: this command is used to disable igmp snooping on the global or VLAN interface.

IGMP snooping is disabled by default on the global or VLAN interfaces.

Notice

Only when IGMP snooping is enabled on the global and VLAN interfaces can the configuration of the other IGMP snooping properties on that interface take effect.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip igmp snooping
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp snooping
```

10.3.2 IGMP Snooping Querier Enablement

【Command】

ip igmp snooping querier

no ip igmp snooping querier

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip igmp snooping querier: this command is used to enable IGMP Snooping querier.

no ip igmp snooping querier: this command is used to disable IGMP Snooping querier.

By default, the IGMP Snooping querier is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp snooping querier
```

10.3.3 IGMP Snooping Port Fast-leave Enablement

【Command】

ip igmp snooping fast-leave

no ip igmp snooping fast-leave

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip igmp-snooping fast-leave: this command is used to enable the fast leave function on all ports of the VLAN interface. Port fast leave means that when the switch receives the IGMP leaving a multicast group message sent by the host from a port, the port is directly deleted from the list of outgoing ports of the corresponding forwarding item.

no ip igmp-snooping fast-leave: this command is used to disable the fast leave function on all ports of the VLAN interface.

By default, the fast leave function of the port is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
```

```
Switch(config-vlanif1)#ip igmp snooping fast-leave
```

10.3.4 IGMP SnoopingPort Suppression Enablement

【Command】

```
ip igmp snooping no-report-suppresstion
no ip igmp snooping no-report-suppresstion
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip igmp snooping no-report-suppresstion: This command is used to disable the port suppression function in all ports of the VLAN interface.

no ip igmp-snooping no-report-suppresstion: This command is used to enable suppression reporting in all ports of the VLAN interface. When the port is in IGMPv1 or IGMPv2, when receiving the leave message, if the port report suppression function is enabled, the report message will not be sent.

Port report suppression is enabled by default.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip igmp snooping no-report-suppresstion
```

10.3.5 Configure the Routing Interface for IGMP Snooping Multicast Group

【Command】

```
ip igmp snooping mrouter inteface <IFNAME>
```

【View】

VLAN-IF Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

IFNAME: port or static aggregate port.

【Description】

ip igmp snooping mrouter inteface: Specifies the multicast group routing port.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)# ip igmp snooping mrouter interface ge1
```

10.3.6 Configure IGMP Snooping Multicast Permanent Group

【Command】

```
ip igmp snooping permanent-group
no ip igmp snooping permanent-group
```

【View】

Global configuration view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

Permanent-group: Configure a multicast group to be a permanent multicast group without aging or leaving.

No IP IgMP Snooping permanent-group: Cancel the permanent multicast group. By default, multicast groups age and leave normally.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)# ip igmp snooping permanent-group
```

10.3.7 Configure IGMP Snooping to Send the Source IP Address

【Command】

ip igmp snooping send source-address A.B.C.D

【View】

Global configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: send the source IP address.

【Description】

ip igmp snooping send sour-address: When there is no IP address in the VLAN, the sending source IP address can be specified. The default IP address is 192.168.0.1.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)# ip igmp snooping send source-address 192.168.0.1
```

10.3.8 Display the IGMP Snooping Multicast Group Routing Interface

【Command】

show ip igmp snooping mrouter inteface <ifname>

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: vlanif interface

【Description】

View the multicast group routing port of the specified interface

【Instance】

```
Switch#show ip igmp snooping mrouter interface vlanif1
VLAN     interface
1        ge2
```

10.3.9 Display IGMP SnoopingMulticast Statistics

【Command】

show ip igmp snooping statistics inteface <IFNAME>

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: vlanif interface

【Description】

View the multicast group statistics of the specified interface

【Instance】

None

10.3.10 Display IGMP Snooping Multicast Group Information

【Command】

show ip igmp snooping groups [interface IFNAME]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: vlanif interface

【Description】

View multicast group information for the current or specified interface of the device.

【Instance】

```
Switch#show ip igmp snooping groups interface ge1 vlan 1
IGMP-Snooping Connected Group Membership, Total is 0
Index Interface   VID   Group            Mac               Mode
Sources
```

11 GMRP and MMRP Configuration

11.1 Overview

The Generic Attribute Registration Protocol (GARP) provides a mechanism for propagating attributes so that a protocol entity can register and deregister attributes. As the carrier of an attribute registration protocol, GVRP can be used to propagate attributes. Mapping the contents of GARP protocol messages to different attributes can support different upper-layer protocol applications.

GMRP (GARP multicast registration protocol) and GVRP (GVRP VLAN registration protocol) are two applications of GVRP, which are used to register and deregister multicast and VLAN attributes respectively. GARP identifies applications through destination MAC addresses. In IEEE Std 802.1D, 01-80-C2-00-00-20 is assigned to multicast application, namely GMRP. IEEE Std 802.1Q assigns 01-80-C2-00-00-21 to the VLAN application (GVRP).

As the carrier of an attribute registration protocol, MRP (Multiple Register Protocol) can be used to propagate attribute messages. The application entities following MRP protocol are called MRP applications, such as MMRP (multiple MAC register protocol) and mvrp (multiple VLAN register protocol). MRP, MVRP and MMRP are upgraded versions of GARP, GVRP and GMRP, which improve the efficiency of attribute declaration and are used to replace GARP, GVRP and GMRP protocols. MMRP/MVRP is used to publish and learn multicast /VLAN configuration information among devices, so that devices can automatically synchronize multicast /VLAN configuration and reduce the configuration work of network managers. After the network topology changes, MMRP/MVRP republishes and learns multicast /VLAN according to the new topology, so as to update synchronously with the network topology in real time.

11.2 Principle Description

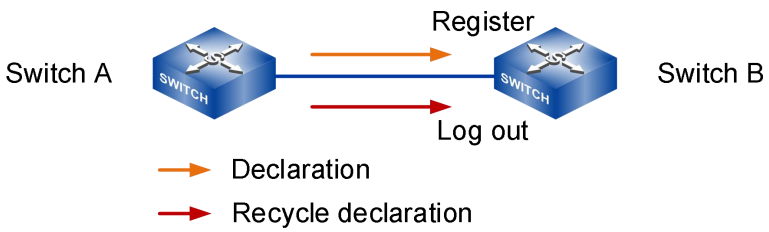
11.2.1 GARP

GARP provides a mechanism for assisting members in the same LAN to distribute, disseminate and register certain information (such as VLAN, multicast address, etc.).

11.2.1.1 GARP implementation mechanism

Every port participating in the protocol on the device can be regarded as an application entity. When GARP application (such as GVRP) is started on the port, the port can be regarded as a GARP application entity.

Through GARP mechanism, the configuration information of a GARP application entity will spread rapidly throughout the whole LAN. As shown in the following figure, the GARP application entity notifies other GARP application entities to register or cancel their own attribute information by sending statements or recycling statements, and registers or cancels each other's attribute information according to the statements or recycling statements sent by other entities.



For example, the GVRP protocol implements VLAN attribute registration and deregistration in the following ways:

- When a port receives a VLAN attribute declaration, the port will register the VLAN attribute contained in the declaration (that is, the port will join the VLAN).
- When a port receives a VLAN attribute reclamation statement, the port will cancel the VLAN attribute contained in the statement (that is, the port exits the VLAN).

11.2.1.2 GARP Message

The exchange of information between GARP application entities is accomplished by transmitting various messages, including Join messages, Leave messages and LeaveAll messages, which cooperate with each other to ensure the registration or

cancellation of information. GVRP is implemented based on GARP, so GVRP also exchanges information through GARP messages.

- Join message
  When a GARP application entity wants other GARP entities to register their own attribute information, it will send a Join message. It also sends Join messages when it receives Join messages from other entities or needs other entities to register because some properties are statically configured. Join messages fall into the following types:
  - JoinEmpty—declare attribute values that it has not registered.
  - JoinIn—declare attribute values that it has registered.
- Leave message
  When a GARP application entity wants other GARP entities to cancel their own attribute information, it will send a Leave message; It also sends Leave messages when it receives Leave messages from other entities and logs off some attributes or logs off some attributes statically. The Leave message is divided into LeaveEmpty and LeaveIn, and the differences between them are as follows:
  - LeaveEmpty: used to unregister an attribute that is not registered.
  - LeaveIn: used to unregister an attribute that has already been registered.
- LeaveAll message
  When each GARP application entity starts, it will start its own LeaveAll timer. When the timer expires, it will send a LeaveAll message to cancel all the attributes, so that other GARP entities can register the attribute information again. It also sends LeaveAll messages when it receives LeaveAll messages from other entities. Send the LeaveAll message and restart the LeaveAll timer to start a new cycle.

11.2.1.3 GARP timer

GARP defines four timers for controlling the sending of various GARP messages.

Notes

- The change of GARP timer value will be applied to all GARP applications (such as GVRP) running in the same LAN.
- Each port of the device maintains its own Hold timer, Join timer and Leave timer independently, while each device maintains only one LeaveAll timer globally.
- The value ranges of Hold timer, Join timer, Leave timer and LeaveAll timer are mutually restricted.

Timer	Lower limit of value	Upper limit of value
Hold timer	10 centiseconds	Less than or equal to half of the Join timer value
Join timer	Greater than or equal to twice the Hold timer value	Less than half of the Leave timer value
Leave timer	Greater than twice the Join timer value	less than the value of the LeaveAll timer
LeaveAll timer	Greater than the Leave timer value on all ports	32765 centiseconds

- Hold timer
  Hold timer is used to control the sending of GARP messages (including Join messages and Leave messages). When the attributes of GARP application entities change or receive GARP messages from other entities, the messages will not be sent out immediately, but all GARP messages to be sent in this period will be packaged into as few messages as possible after the Hold timer expires, thus reducing the number of messages sent and saving bandwidth resources.
- Join timer
  Join timer is used to control the sending of Join message. In order to ensure that the Join message can be reliably transmitted to other entities, the GARP application entity will wait for a Join timer interval after sending out the Join message: if it receives the JoinIn message sent by other entities before the timer expires, it will not resend the Join message; Otherwise, it will resend the Join message once.

  Note
  Not every attribute has its own Join timer, but every GARP application entity shares one. Therefore, the Join timer should be large enough to ensure that all attributes can be sent out in one declaration process.

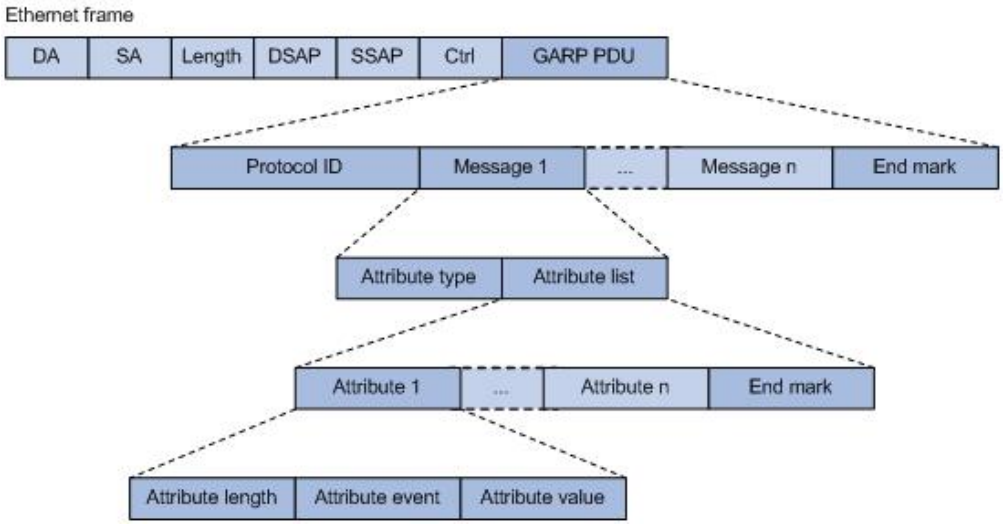
- Leave timer
  The Leave timer controls the deregistration of attributes. When an GARP participant wishes other participants to deregister its attributes, it sends a Leave message. On receiving a Leave message, MRP starts the Leave timer, and deregisters the attributes if it does not receive any Join message for the attributes before the Leave timer expires.
- LeaveAll timer
  Every GARP application entity starts its own LeaveAll timer when it starts. when the timer expires, the GARP application entity will send a LeaveAll message to

the outside, so that other entities can re-register the attribute information. When the LeaveAll timer expires, MRP sends out a LeaveAll message and restarts the LeaveAll timer. The entity receiving the LeaveAll message will restart all timers, including the LeaveAll timer.

Notes

- Every time the LeaveAll timer times out, it will cause all attributes of the whole network to be logged off. Because it has a wide range of influences, the value of the LeaveAll timer should not be too small, and it must be greater than the value of the LeaveAll timer on all ports. It is suggested that the value of the Leave All timer configured by the user should not be less than its default value (i.e. 1000 centiseconds).
- Although the values of the LeaveAll timer may be different on all devices in the whole network, these devices will send the LeaveAll message with the minimum value of the LeaveAll timer as the cycle. This is because every device will clear its own LeaveAll timer after receiving the LeaveAll message, and only the device with the minimum LeaveAll timer value can send out the LeaveAll message in time, so in fact, only the LeaveAll timer with the minimum value of the whole network will take effect.

11.2.1.4 Packet encapsulation format of GARP protocol



As shown in the above figure, GARP protocol message adopts IEEE 802.3 Ethernet encapsulation format, and the description of main fields is shown in the following table.

Field	Note
GARP PDU	GARP PDU (protocol data unit) encapsulated in GARP protocol message
Protocol ID	Protocol number, the protocol number of GARP PDU is 0x0001

Field	Note
Message	Attribute messages, each message consists of two fields, Attribute type and Attribute list
End Mark	End mark, with a value of 0x00
Attribute type	Attribute types are defined by specific GARP applications. A value of 0x01 indicates VLAN ID, representing GVRP application
Attribute list	Attribute list, which comprises multiple attributes.
Attribute	Attributes, each of which consists of three fields: Attribute length, Attribute event and Attribute value
Attribute length	The attribute length (including this field) ranges from 2 to 255 in bytes
Attribute event	The value and meaning of the event described by the attribute are as follows: • 0x00: indicates the LeaveAll event • 0x01: indicates JoinEmpty event • 0x02: indicates JoinIn event • 0x03: indicates the LeaveEmpty event • 0x04: indicates the LeaveIn event • 0x05: indicates the Empty event
Attribute value	Attribute value. The value of attribute applied by GVRP is VLAN ID, but when the value of Attribute event field is 0x00 (i.e. LeaveAll event), this field is invalid.

GARP protocol messages take specific multicast MAC addresses as destination MAC, for example, the destination MAC address of GVRP is 01-80-C2-00-00-21, and the destination MAC address of GMRP is 01-80-C2-00-00-20. After receiving the message from GARP application entity, the device will distribute it to different GARP applications for processing according to its destination MAC address.

11.2.1.5 GMRP Implementation

GMRP is a multicast registration protocol based on GARP, which is used to maintain multicast registration information in switches. All switches supporting GMRP can receive multicast registration information from other switches, and dynamically update local multicast registration information, and can also spread local multicast registration information to other switches. This information exchange mechanism

ensures the consistency of multicast information maintained by all GMRP-enabled devices in the same switching network.

GMRP defines two attribute types:

- Multicast member information
- Multicast service demand information

When a host wants to join a multicast group, it will send out GMRP join message. The switch adds the port receiving the GMRP join message to the multicast group, and broadcasts the GMRP join message in the VLAN where the receiving port is located, so that the multicast source in the VLAN can know the existence of multicast members. When a multicast source sends a multicast message to a multicast group, the switch only forwards the multicast message to the port connected with the members of the multicast group, thus realizing layer 2 multicast in VLAN.

We call the multicast added manually static multicast, and the multicast created by GMRP protocol dynamic multicast. GMRP has three registration modes, and different registration modes deal with static multicast and dynamic multicast differently.

- Normal mode
  Dynamic multicast is allowed to register on the port, and the announcement messages of static multicast and dynamic multicast are sent at the same time.
- Fixed mode
  Dynamic multicast is not allowed to be registered on the port, and the registered multicast group is reserved.
- Forbidden mode
  Dynamic multicast is not allowed to register on the port, and remains unregistered.

11.2.2 MRP

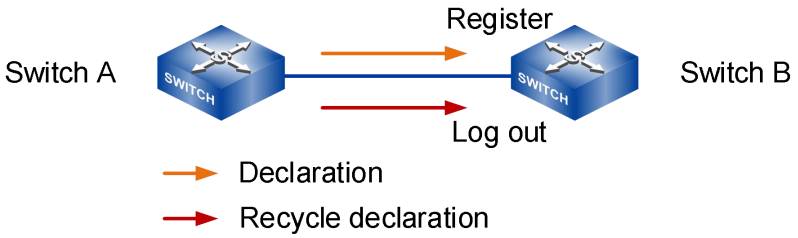
Different from GARP, MRP allows participants in the same LAN to declare, propagate, and register information (for example, VLAN information) on a per Multiple Spanning Tree Instance (MSTI) basis.

11.2.2.1 MRP implementation mechanism

Each port that participates in an MRP application (for example, MVRP) is called an “MRP participant”.

MRP rapidly propagates the configuration information of an MRP participant throughout the LAN. As shown in the following figure, MRP application entities notify other MRP application entities to register or cancel their own attribute information by

sending two protocol messages: declaration or recovery declaration, and register or cancel each other's attribute information according to the declaration or recovery declaration sent by other MRP entities.



11.2.2.2 MRP Message

MRP exchanges information among MRP participants by advertising MRP messages, including Join, New, Leave, and LeaveAll. Join and New messages are declarations, and Leave and LeaveAll messages are withdrawals. As an MRP application, MVRP also uses MRP messages for information exchange.

- Join message

  An MRP participant sends Join messages when it wishes to declare its attribute values and receives Join messages from other MRP participants. When receiving a Join message, an MRP participant sends a Join message to all participants except the sender. Join messages fall into the following types:

  - JoinEmpty—declare attribute values that it has not registered.
  - JoinIn—declare attribute values that it has registered.

- New message

  When the Multiple Spanning Tree Protocol (MSTP) topology changes, in other words, when an MSTP TcDetected event occurs, an MRP participant sends New messages to declare the topology change. On receiving a New message, an MRP participant sends a New message out each port except the receiving port. Similar to a Join message, a New message enables MRP participants to register attributes.

- Leave message

  An MRP participant sends Leave messages when it wishes to withdraw declarations of its attribute values and receives Leave messages from other participants. When receiving a Leave message, an MRP participant sends a Leave message to all participants except the sender.

- LeaveAll message

  Each MRP participant is configured with an individual LeaveAll timer. When the timer expires, the MRP participant sends LeaveAll messages to deregister all

attributes, so that any other MRP participant can re-register all attributes. This process periodically clears the useless attributes in the network. On receiving a LeaveAll message, MRP determines whether to send a Join message to request the sender to re-register these attributes according to attribute status. On sending a LeaveAll message, MRP restarts the LeaveAll timer.

11.2.2.3 MRP timer

The implementation of MRP uses the following timers to control MRP message transmission.

- Periodic timer
  On startup, an MRP participant starts its own Periodic timer to control MRP message transmission. The MRP participant collects the MRP messages to be sent before the Periodic timer expires, and sends the MRP messages in as few packets as possible when the Periodic timer expires and meanwhile restarts the Periodic timer. This mechanism reduces the number of MRP protocol packets periodically sent.

  Notes
  You can enable or disable the Periodic timer at the CLI. When you disable the Periodic timer, MRP will not send MRP messages.

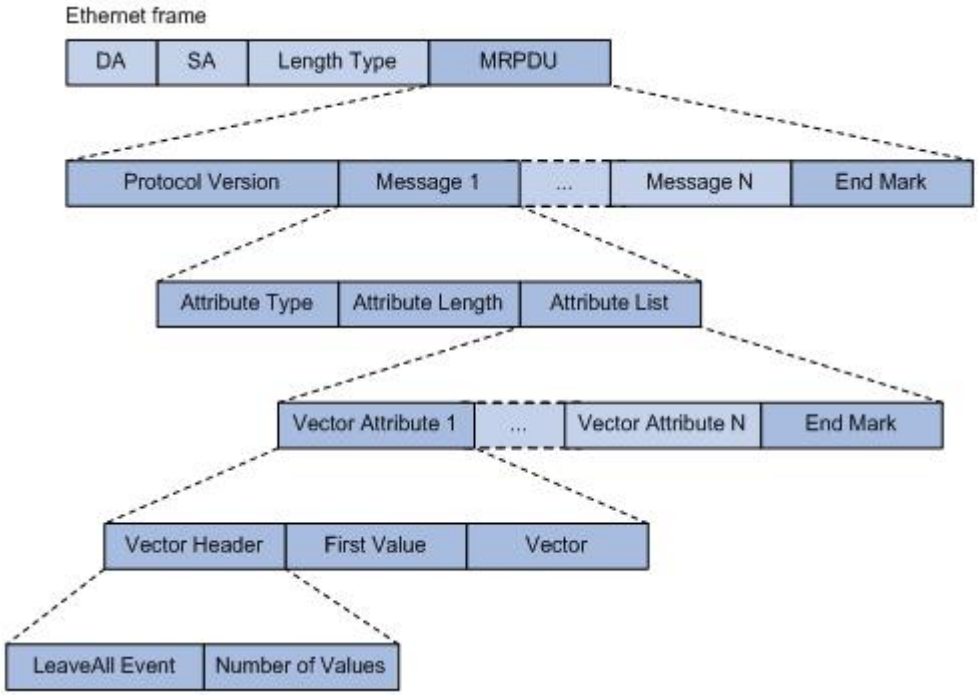
- Join timer
  The Join timer control the transmission of Join messages. To make sure that Join messages can be reliably transmitted to other participants, an MRP participant waits for a period of the Join timer after sending a Join message. If the participant receives JoinIn messages from other participants before the Join timer expires, the participant does not re-send the Join message. When both the Join timer and the Periodic timer expire, the participant re-sends the Join message.
- Leave timer
  The Leave timer controls the deregistration of attributes. When an MRP participant wishes other participants to deregister its attributes, it sends a Leave message. On receiving a Leave message, MRP starts the Leave timer, and deregisters the attributes if it does not receive any Join message for the attributes before the Leave timer expires.
- LeaveAll timer
  On startup, an MRP participant starts its own LeaveAll timer. When the LeaveAll timer expires, MRP sends out a LeaveAll message and restarts the LeaveAll timer. On receiving the LeaveAll message, other participants re-register all the

attributes and re-start their LeaveAll timer.

Notes

Though MRP participants throughout the network may be configured with different LeaveAll timers, an MRP participant sends LeaveAll messages at the smallest interval among the neighboring participants' LeaveAll timers. At the next startup, the LeaveAll timer of each participant randomly changes within a certain range.

11.2.2.4 Message encapsulation format of MRP protocol



As shown in the above figure, GARP protocol message adopts IEEE 802.3 Ethernet encapsulation format, and the description of main fields is shown in the following table.

Field	Note
MRPDU	MRP protocol data unit (MRPDU) encapsulated in the MRP protocol packet.
Protocol Version	Protocol version, which is 0.
Message	Attribute message, which comprises the Attribute Type, Attribute Length, and Attribute List fields.
End Mark	End mark of the MRPDU or an attribute list field. This field is fixed at 0x00.
Attribute Type	Attribute type, which is VID Vector specified by the value of

Field	Note
	1.
Attribute Length	Length of the FirstValue field, which is 2 as specified by MVRP.
Attribute List	Attribute list, which comprises multiple attributes.
Vector Attribute	Vector attribute, which comprises the VectorHeader, FirstValue, and Vector fields.
Vector Header	Vector header, which comprises the LeaveAllEvent and NumberOfValues fields.
FirstValue	The first attribute value encapsulated in the MVRP protocol packet.
Vector	Attribute events, where each byte specifies three attribute events. The attribute events include: • 0x00: New operator • 0x01: JoinIn operator • 0x02: In operator • 0x03: JoinMt operator • 0x04: Mt operator • 0x05: Lv operator Assume that the three attribute events sharing a byte are A1, A2, and A3. The value of the byte A1A2A3 is ((A1 * 6 + A2) * 6) + A3, which ranges from 0 to 255.
LeaveAll Event	LeaveAll event indicator: • 0—Not a LeaveAll event • 1—A LeaveAll event
Number of Values	A 13-bit field, which shows the number of attribute values encoded in the Vector field.

MRP protocol message takes the specific multicast MAC address as the destination MAC. For example, the destination MAC address of MVRP is 01-80-C2-00-21, and the Type is 88F5. The destination MAC address of MMRP is 01-80-C2-00-00-20, and Type 88F6. When a device receives a packet from an MRP participant, it delivers the packet to the MRP application identified by the destination MAC address.

11.2.2.5 Implementation of MMRP

MMRP is a kind of MRP application, which maintains the multicast dynamic registration information in equipment based on MRP working mechanism, and

spreads the information to other equipment: after the equipment starts MMRP, it can receive the multicast registration information from other equipment and dynamically update the local multicast registration information, including the current multicast members and which ports these multicast members can reach; In addition, the device can also spread the local multicast registration information to other devices, so that the multicast information of all devices in the same LAN can be consistent.

The multicast registration information propagated by MMRP includes both static registration information manually configured locally and dynamic registration information from other devices.

The MMRP protocol implements multicast attribute registration and deregistration as follows:

- When a port receives a multicast attribute declaration, the port will register the multicast attribute contained in the declaration (that is, the port joins the multicast).
- When a port receives a multicast attribute recovery statement, the port will cancel the multicast attribute contained in the statement (that is, the port quits the multicast).

We call the multicast added manually static multicast, and the multicast created by MMRP protocol dynamic multicast. MMRP has three registration modes, and different registration modes deal with static multicast and dynamic multicast differently.

- Normal mode
  The interface in this mode allows dynamic multicast registration or deregistration, and allows dynamic and static multicast declarations to be sent.
- Fixed mode
  In this mode, the interface prohibits the cancellation of dynamic multicast, but allows the sending of dynamic and static multicast statements, and the received MMRP messages will be ignored and discarded. That is to say, the learned dynamic multicast will not be cancelled, and the new dynamic multicast will not be learned at the same time.
- Forbidden mode
  In this mode, the interface prohibits the registration of dynamic multicast, but allows the sending of statements of dynamic and static multicast, and the received MMRP messages will be ignored and discarded. That is to say, registration of dynamic multicast is not allowed, and once the learned dynamic multicast is cancelled, it will not be re-learned.

11.3 Configure GMRP or MMRP

11.3.1 GlobalGMRP or MMRP Enablement

【Command】

(gmrp | mmrp) (enable | disable)

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

(gmrp | mmrp) enable: this command is used to enable the global GMRP (MMRP) function.

(gmrp | mmrp) disable: this command is used to disable the global GMRP（MMRP）function.

By default, the global GMRP（MMRP） is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#gmrp enable
```

11.3.2 Port GMRP or MMRP Enablement

【Command】

(gmrp | mmrp) (enable | disable)

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

port (gmrp | mmrp) enable: this command is used to enable the GMRP (MMRP) function of the port.

port (gmrp | mmrp) disable: this command is used to disable the GMRP（MMRP）function of the port.

By default, the GMRP（MMRP）function of the port is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#gmrp enable
Switch(config)#interface ge5
Switch(config-ge5)#gmrp enable
```

11.3.3 GMRP or MMRP Registration Mode

【Command】

(gmrp | mmrp) registration (fixed| forbidden | normal | restricted)

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

fixed：Fixed mode.

forbidden：Forbidden mode.

normal: Normal mode, allowing registering and deregistering multicast dynamically.

restricted：Restricted mode.

【Description】

(gmrp | mmrp) registration: this command is used for GMRP port registration mode.

By default, the GMRP port registration mode is normal.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#gmrp enable
Switch(config)#interface ge5
Switch(config-ge5)#gmrp registration normal
```

11.3.4 GMRP or MMRP Timer

【Command】

(gmrp | mmrp) timer (join| leave| leaveall) <TIMER_VALUE>

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

leaveall | join | leave: represent leave All, join and leave three timers respectively. After each port starts GMRP/MMRP, it starts the LeaveAll timer at the same time, and the port will send the LeaveAll message circularly to make other ports re-register all their attribute information. GMRP/MMRP port can send each Join packet out twice to ensure the reliable transmission of messages, and the time interval between the two transmissions is controlled by the Join timer. GMRP/MMRP port can send each Join packet out twice to ensure the reliable transmission of messages, and the time interval between the two transmissions is controlled by the Join timer.

<TIMER_VALUE> : timer value, leave All defaults to 1000; The default value for join is 20; The default value for leave is 60. Unit: centiseconds.

【Description】

(gmrp | mmrp) timer: this command is used to configure the GMRP/MMRP port leave All, join and leave timers.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#gmrp enable
Switch(config)#interface ge5
Switch(config-ge5)#gmrp timer leave 100
```

11.3.5 Display GMRP or MMRP Configuration Information

【Command】

show (gmrp | mmrp) configuration

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show gmrp | mmrp configuration: this command is used to display GMRP| MMRP configuration information.

【Instance】

```
*Switch#show gmrp configuration
```

11.3.6 Display GMRP or MMRP State Machine Information

【Command】

show (gmrp | mmrp) machine

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show gmrp | mmrp machine command: this command is used to display GMRP| MMRP state machine information.

【Instance】

```
Switch> enable
Switch#show gmrp machine
```

11.3.7 Display GMRP or MMRP Message Statistics

【Command】

show (gmrp | mmrp) statistics vlanid [<VLANID>]

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

vlanid：VLAN ID.

【Description】

show gmrp | mmrp statistics: this command is used to display GMRP| MMRP message statistics.

【Instance】

```
Switch> enable
Switch#show gmrp statistics vlanid 3
```

11.3.8 Display GMRP or MMRP Timer Information

【Command】

show (gmrp | mmrp) timer <IFNAME>

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

Ifname：port name.

【Description】

show gmrp | mmrp timer: this command is used to display the GMRP| MMRP port timer information.

【Instance】

```
Switch> enable
Switch#show gmrp timer ge2
```

12 GVRP and MVRP Configuration

12.1 Overview

The Generic Attribute Registration Protocol (GARP) provides a mechanism for propagating attributes so that a protocol entity can register and deregister attributes. As the carrier of an attribute registration protocol, GVRP can be used to propagate attributes. Mapping the contents of GARP protocol messages to different attributes can support different upper-layer protocol applications.

GMRP (GARP multicast registration protocol) and GVRP (GVRP VLAN registration protocol) are two applications of GVRP, which are used to register and deregister multicast and VLAN attributes respectively. GARP identifies applications through destination MAC addresses. In IEEE Std 802.1D, 01-80-C2-00-00-20 is assigned to multicast application, namely GMRP. IEEE Std 802.1Q assigns 01-80-C2-00-00-21 to the VLAN application (GVRP).

As the carrier of an attribute registration protocol, MRP (Multiple Register Protocol) can be used to propagate attribute messages. The application entities following MRP protocol are called MRP applications, such as MMRP (multiple MAC register protocol) and mvrp (multiple VLAN register protocol). MRP, MVRP and MMRP are upgraded versions of GARP, GVRP and GMRP, which improve the efficiency of attribute declaration and are used to replace GARP, GVRP and GMRP protocols. MMRP/MVRP is used to publish and learn multicast /VLAN configuration information among devices, so that devices can automatically synchronize multicast /VLAN configuration and reduce the configuration work of network managers. After the network topology changes, MMRP/MVRP republishes and learns multicast /VLAN according to the new topology, so as to update synchronously with the network topology in real time.

12.2 Principle Description

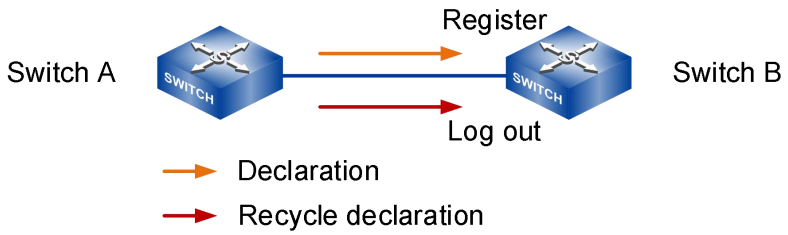
12.2.1 GARP

GARP provides a mechanism for assisting members in the same LAN to distribute, disseminate and register certain information (such as VLAN, multicast address, etc.).

12.2.1.1 GARP implementation mechanism

Every port participating in the protocol on the device can be regarded as an application entity. When GARP application (such as GVRP) is started on the port, the port can be regarded as a GARP application entity.

Through GARP mechanism, the configuration information of a GARP application entity will spread rapidly throughout the whole LAN. As shown in the following figure, the GARP application entity notifies other GARP application entities to register or cancel their own attribute information by sending statements or recycling statements, and registers or cancels each other's attribute information according to the statements or recycling statements sent by other entities.



For example, the GVRP protocol implements VLAN attribute registration and deregistration in the following ways:

- When a port receives a VLAN attribute declaration, the port will register the VLAN attribute contained in the declaration (that is, the port will join the VLAN).
- When a port receives a VLAN attribute reclamation statement, the port will cancel the VLAN attribute contained in the statement (that is, the port exits the VLAN).

12.2.1.2 GARP Message

The exchange of information between GARP application entities is accomplished by transmitting various messages, including Join messages, Leave messages and LeaveAll messages, which cooperate with each other to ensure the registration or

cancellation of information. GVRP is implemented based on GARP, so GVRP also exchanges information through GARP messages.

- Join message
  When a GARP application entity wants other GARP entities to register their own attribute information, it will send a Join message. It also sends Join messages when it receives Join messages from other entities or needs other entities to register because some properties are statically configured. Join messages fall into the following types:
  - JoinEmpty—declare attribute values that it has not registered.
  - JoinIn—declare attribute values that it has registered.
- Leave message
  When a GARP application entity wants other GARP entities to cancel their own attribute information, it will send a Leave message; It also sends Leave messages when it receives Leave messages from other entities and logs off some attributes or logs off some attributes statically. The Leave message is divided into LeaveEmpty and LeaveIn, and the differences between them are as follows:
  - LeaveEmpty: used to unregister an attribute that is not registered.
  - LeaveIn: used to unregister an attribute that has already been registered.
- LeaveAll message
  When each GARP application entity starts, it will start its own LeaveAll timer. When the timer expires, it will send a LeaveAll message to cancel all the attributes, so that other GARP entities can register the attribute information again. It also sends LeaveAll messages when it receives LeaveAll messages from other entities. Send the LeaveAll message and restart the LeaveAll timer to start a new cycle.

12.2.1.3 GARP timer

GARP defines four timers for controlling the sending of various GARP messages.

Notes

- The change of GARP timer value will be applied to all GARP applications (such as GVRP) running in the same LAN.
- Each port of the device maintains its own Hold timer, Join timer and Leave timer independently, while each device maintains only one LeaveAll timer globally.
- The value ranges of Hold timer, Join timer, Leave timer and LeaveAll timer are mutually restricted.

Timer	Lower limit of value	Upper limit of value
Hold timer	10 centiseconds	Less than or equal to half of the Join timer value
Join timer	Greater than or equal to twice the Hold timer value	Less than half of the Leave timer value
Leave timer	Greater than twice the Join timer value	less than the value of the LeaveAll timer
LeaveAll timer	Greater than the Leave timer value on all ports	32765 centiseconds

- Hold timer
  Hold timer is used to control the sending of GARP messages (including Join messages and Leave messages). When the attributes of GARP application entities change or receive GARP messages from other entities, the messages will not be sent out immediately, but all GARP messages to be sent in this period will be packaged into as few messages as possible after the Hold timer expires, thus reducing the number of messages sent and saving bandwidth resources.
- Join timer
  Join timer is used to control the sending of Join message. In order to ensure that the Join message can be reliably transmitted to other entities, the GARP application entity will wait for a Join timer interval after sending out the Join message: if it receives the JoinIn message sent by other entities before the timer expires, it will not resend the Join message; Otherwise, it will resend the Join message once.

  Note

  Not every attribute has its own Join timer, but every GARP application entity shares one. Therefore, the Join timer should be large enough to ensure that all attributes can be sent out in one declaration process.

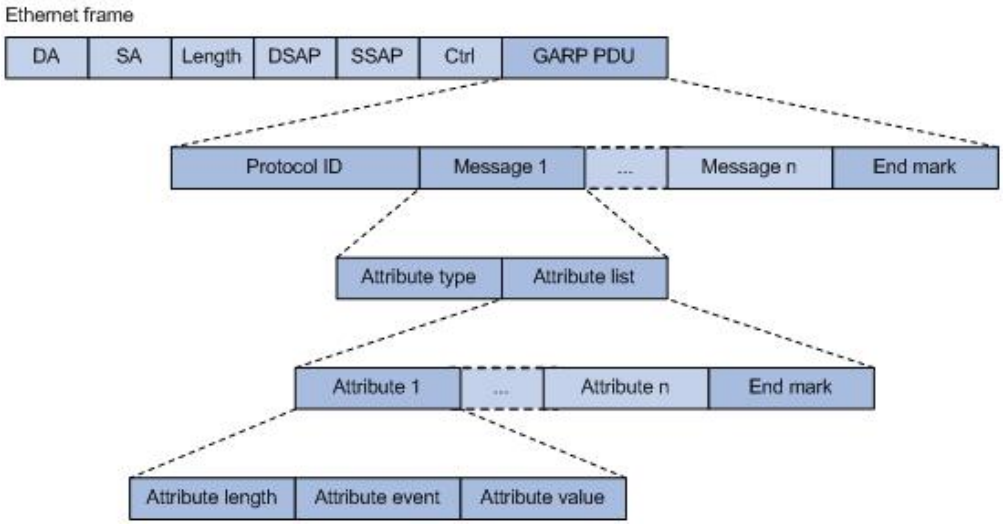
- Leave timer
  The Leave timer controls the deregistration of attributes. When an GARP participant wishes other participants to deregister its attributes, it sends a Leave message. On receiving a Leave message, MRP starts the Leave timer, and deregisters the attributes if it does not receive any Join message for the attributes before the Leave timer expires.
- LeaveAll timer
  Every GARP application entity starts its own LeaveAll timer when it starts. when the timer expires, the GARP application entity will send a LeaveAll message to

the outside, so that other entities can re-register the attribute information. When the LeaveAll timer expires, MRP sends out a LeaveAll message and restarts the LeaveAll timer. The entity receiving the LeaveAll message will restart all timers, including the LeaveAll timer.

Notes

- Every time the LeaveAll timer times out, it will cause all attributes of the whole network to be logged off. Because it has a wide range of influences, the value of the LeaveAll timer should not be too small, and it must be greater than the value of the LeaveAll timer on all ports. It is suggested that the value of the Leave All timer configured by the user should not be less than its default value (i.e. 1000 centiseconds).
- Although the values of the LeaveAll timer may be different on all devices in the whole network, these devices will send the LeaveAll message with the minimum value of the LeaveAll timer as the cycle. This is because every device will clear its own LeaveAll timer after receiving the LeaveAll message, and only the device with the minimum LeaveAll timer value can send out the LeaveAll message in time, so in fact, only the LeaveAll timer with the minimum value of the whole network will take effect.

12.2.1.4 Packet encapsulation format of GARP protocol



As shown in the above figure, GARP protocol message adopts IEEE 802.3 Ethernet encapsulation format, and the description of main fields is shown in the following table.

Field	Note
GARP PDU	GARP PDU (protocol data unit) encapsulated in GARP protocol message
Protocol ID	Protocol number, the protocol number of GARP PDU is 0x0001

Field	Note
Message	Attribute messages, each message consists of two fields, Attribute type and Attribute list
End Mark	End mark, with a value of 0x00
Attribute type	Attribute types are defined by specific GARP applications. A value of 0x01 indicates VLAN ID, representing GVRP application
Attribute list	Attribute list, which comprises multiple attributes.
Attribute	Attributes, each of which consists of three fields: Attribute length, Attribute event and Attribute value
Attribute length	The attribute length (including this field) ranges from 2 to 255 in bytes
Attribute event	The value and meaning of the event described by the attribute are as follows: • 0x00: indicates the LeaveAll event • 0x01: indicates JoinEmpty event • 0x02: indicates JoinIn event • 0x03: indicates the LeaveEmpty event • 0x04: indicates the LeaveIn event • 0x05: indicates the Empty event
Attribute value	Attribute value. The value of attribute applied by GVRP is VLAN ID, but when the value of Attribute event field is 0x00 (i.e. LeaveAll event), this field is invalid.

GARP protocol messages take specific multicast MAC addresses as destination MAC, for example, the destination MAC address of GVRP is 01-80-C2-00-00-21, and the destination MAC address of GMRP is 01-80-C2-00-00-20. After receiving the message from GARP application entity, the device will distribute it to different GARP applications for processing according to its destination MAC address.

12.2.1.5 GVRP Implementation

As an GARP application, GVRP uses the operating mechanism of GARP to maintain and propagate dynamic VLAN registration information throughout the network. In a LAN, each GVRP-enabled device can receive the VLAN registration information from other MVRP devices, and dynamically update its local database, including active VLANs and the ports through which a VLAN can be reached. This makes sure that all MVRP-enabled devices in a LAN maintain the same VLAN information.

The VLAN information propagated by GVRP includes not only locally, manually configured static VLAN information but also dynamic VLAN information from other devices.

VLANs created manually, locally are called “static VLANs”, and VLANs learned through GVRP are called “dynamic VLANs”. The following GVRP registration modes are available.

- Normal mode
  A port in normal registration mode performs dynamic VLAN registrations and deregistrations, and sends declarations for dynamic and static VLANs.
- Fixed mode
  Ports in this mode are forbidden to register or deregister dynamic VLANs, and only static VLAN declarations can be sent. That is to say, even if all VLANs are allowed to pass through the Trunk port in this mode, the VLAN actually passed through can only be the VLAN created manually.
- Forbidden mode
  Ports in this mode are forbidden to register or deregister dynamic VLANs, and only VLAN 1 declarations can be sent. That is to say, even if all VLANs are allowed to pass through the Trunk port in this mode, the VLAN actually passed through can only be VLAN 1.

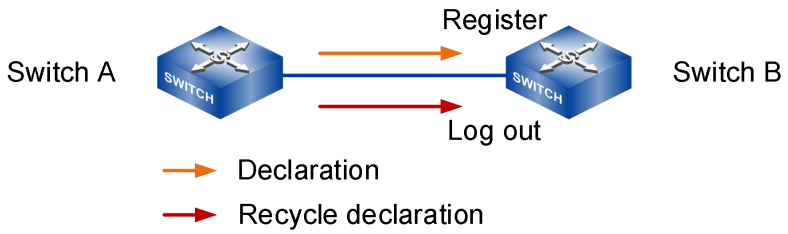
12.2.2 MRP

Different from GARP, MRP allows participants in the same LAN to declare, propagate, and register information (for example, VLAN information) on a per Multiple Spanning Tree Instance (MSTI) basis.

12.2.2.1 MRP implementation mechanism

Each port that participates in an MRP application (for example, MVRP) is called an “MRP participant”.

MRP rapidly propagates the configuration information of an MRP participant throughout the LAN. As shown in the following figure, MRP application entities notify other MRP application entities to register or cancel their own attribute information by sending two protocol messages: declaration or recovery declaration, and register or cancel each other's attribute information according to the declaration or recovery declaration sent by other MRP entities.



12.2.2.2 MRP Message

MRP exchanges information among MRP participants by advertising MRP messages, including Join, New, Leave, and LeaveAll. Join and New messages are declarations, and Leave and LeaveAll messages are withdrawals. As an MRP application, MVRP also uses MRP messages for information exchange.

- Join message

  An MRP participant sends Join messages when it wishes to declare its attribute values and receives Join messages from other MRP participants. When receiving a Join message, an MRP participant sends a Join message to all participants except the sender. Join messages fall into the following types:

  - JoinEmpty—declare attribute values that it has not registered.
  - JoinIn—declare attribute values that it has registered.

- New message

  When the Multiple Spanning Tree Protocol (MSTP) topology changes, in other words, when an MSTP TcDetected event occurs, an MRP participant sends New messages to declare the topology change. On receiving a New message, an MRP participant sends a New message out each port except the receiving port. Similar to a Join message, a New message enables MRP participants to register attributes.

- Leave message

  An MRP participant sends Leave messages when it wishes to withdraw declarations of its attribute values and receives Leave messages from other participants. When receiving a Leave message, an MRP participant sends a Leave message to all participants except the sender.

- LeaveAll message

  Each MRP participant is configured with an individual LeaveAll timer. When the timer expires, the MRP participant sends LeaveAll messages to deregister all attributes, so that any other MRP participant can re-register all attributes. This process periodically clears the useless attributes in the network. On receiving a LeaveAll message, MRP determines whether to send a Join message to request

the sender to re-register these attributes according to attribute status. On sending a LeaveAll message, MRP restarts the LeaveAll timer.

12.2.2.3 MRP timer

The implementation of MRP uses the following timers to control MRP message transmission.

- Periodic timer
  On startup, an MRP participant starts its own Periodic timer to control MRP message transmission. The MRP participant collects the MRP messages to be sent before the Periodic timer expires, and sends the MRP messages in as few packets as possible when the Periodic timer expires and meanwhile restarts the Periodic timer. This mechanism reduces the number of MRP protocol packets periodically sent.

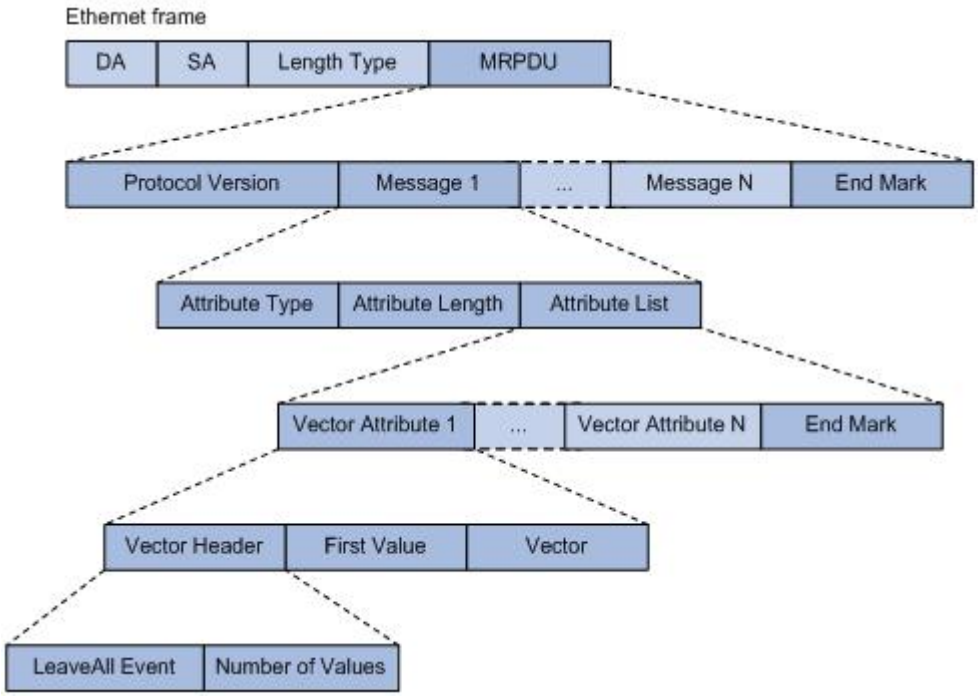
  Notes
  You can enable or disable the Periodic timer at the CLI. When you disable the Periodic timer, MRP will not send MRP messages.

- Join timer
  The Join timer control the transmission of Join messages. To make sure that Join messages can be reliably transmitted to other participants, an MRP participant waits for a period of the Join timer after sending a Join message. If the participant receives JoinIn messages from other participants before the Join timer expires, the participant does not re-send the Join message. When both the Join timer and the Periodic timer expire, the participant re-sends the Join message.
- Leave timer
  The Leave timer controls the deregistration of attributes. When an MRP participant wishes other participants to deregister its attributes, it sends a Leave message. On receiving a Leave message, MRP starts the Leave timer, and deregisters the attributes if it does not receive any Join message for the attributes before the Leave timer expires.
- LeaveAll timer
  On startup, an MRP participant starts its own LeaveAll timer. When the LeaveAll timer expires, MRP sends out a LeaveAll message and restarts the LeaveAll timer. On receiving the LeaveAll message, other participants re-register all the attributes and re-start their LeaveAll timer.

  Notes

Though MRP participants throughout the network may be configured with different LeaveAll timers, an MRP participant sends LeaveAll messages at the smallest interval among the neighboring participants' LeaveAll timers. At the next startup, the LeaveAll timer of each participant randomly changes within a certain range.

12.2.2.4 Message encapsulation format of MRP protocol



As shown in the above figure, GARP protocol message adopts IEEE 802.3 Ethernet encapsulation format, and the description of main fields is shown in the following table.

Field	Note
MRPDU	MRP protocol data unit (MRPDU) encapsulated in the MRP protocol packet.
Protocol Version	Protocol version, which is 0.
Message	Attribute message, which comprises the Attribute Type, Attribute Length, and Attribute List fields.
End Mark	End mark of the MRPDU or an attribute list field. This field is fixed at 0x00.
Attribute Type	Attribute type, which is VID Vector specified by the value of 1.
Attribute Length	Length of the FirstValue field, which is 2 as specified by MVRP.

Field	Note
Attribute List	Attribute list, which comprises multiple attributes.
Vector Attribute	Vector attribute, which comprises the VectorHeader, FirstValue, and Vector fields.
Vector Header	Vector header, which comprises the LeaveAllEvent and NumberOfValues fields.
FirstValue	The first attribute value encapsulated in the MVRP protocol packet.
Vector	Attribute events, where each byte specifies three attribute events. The attribute events include: • 0x00: New operator • 0x01: JoinIn operator • 0x02: In operator • 0x03: JoinMt operator • 0x04: Mt operator • 0x05: Lv operator Assume that the three attribute events sharing a byte are A1, A2, and A3. The value of the byte A1A2A3 is ((A1 * 6 + A2) * 6) + A3, which ranges from 0 to 255.
LeaveAll Event	LeaveAll event indicator: • 0—Not a LeaveAll event • 1—A LeaveAll event
Number of Values	A 13-bit field, which shows the number of attribute values encoded in the Vector field.

MRP protocol message takes the specific multicast MAC address as the destination MAC. For example, the destination MAC address of MVRP is 01-80-C2-00-21, and the Type is 88F5. The destination MAC address of MMRP is 01-80-C2-00-00-20, and Type 88F6. When a device receives a packet from an MRP participant, it delivers the packet to the MRP application identified by the destination MAC address.

12.2.2.5 Implementation of MVRP

As an MRP application, MVRP uses the operating mechanism of MRP to maintain and propagate dynamic VLAN registration information throughout the network. In a LAN, each MVRP-enabled device can receive the VLAN registration information from other MVRP devices, and dynamically update its local database, including active

VLANs and the ports through which a VLAN can be reached. This makes sure that all MVRP-enabled devices in a LAN maintain the same VLAN information.

The VLAN information propagated by MVRP includes not only locally, manually configured static VLAN information but also dynamic VLAN information from other devices.

MVRP registers and deregisters VLAN attributes as follows:

- When a port receives a VLAN attribute declaration, the port will register the VLAN attribute contained in the declaration (that is, the port will join the VLAN).
- When a port receives a VLAN attribute reclamation statement, the port will cancel the VLAN attribute contained in the statement (that is, the port exits the VLAN).

VLANs created manually, locally are called “static VLANs”, and VLANs learned through MVRP are called “dynamic VLANs”. The following MVRP registration modes are available.

- Normal mode
  A port in normal registration mode performs dynamic VLAN registrations and deregistrations, and sends declarations and withdrawals for dynamic and static VLANs.
- Fixed mode
  A port in fixed registration mode disables deregistering dynamic VLANs, sends declarations for dynamic VLANs and static VLANs, and drops received MVRP protocol packets. As a result, a trunk port in fixed registration mode does not deregister or register dynamic VLANs.
- Forbidden mode
  A port in forbidden registration mode disables registering dynamic VLANs, sends declarations for dynamic VLANs and static VLANs, and drops received MVRP protocol packets. As a result, a trunk port in forbidden registration mode does not register dynamic VLANs, and does not re-register a dynamic VLAN when the VLAN is deregistered.

12.3 Configure GVRP and MVRP

12.3.1 Global GVRP or MVRP Enablement

【Command】

```
(gvrp | mvrp) (enable | disable)
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

(gvrp | mvrp) enable: this command is used to enable the global GVRP (MVRP) function.

(gvrp | mvrp) disable: this command is used to disable the global GVRP (MVRP) function.

By default, the global GVRP (MVRP) function is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#gvrp enable
```

12.3.2 GVRP or MVRP Dynamic VLAN Enablement

【Command】

(gvrp | mvrp) dynamic-vlan-creation (enable | disable)

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

(gvrp | mvrp) dynamic-vlan-creation enable: this command is used to enable the dynamic creation of VLAN functions.

(gvrp | mvrp) dynamic-vlan-creation disable: this command is used to disable the dynamic creation of VLAN functions.

By default, the dynamic creation VLAN feature is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#gvrp enable
```

```
Switch(config)#gvrp dynamic-vlan-creation enable
```

12.3.3 Port GVRP or MVRP Enablement

【Command】

```
(gvrp | mvrp) (enable | disable)
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

(gvrp | mvrp) enable: this command is used to enable port GVRP (MVRP) function.

(gvrp | mvrp) disable: this command is used to disable the GVRP（MVRP） function.

By default, the GVRP（MVRP）function of the port is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#gvrp enable
Switch(config)#gvrp dynamic-vlan-creation enable
Switch(config)#interface ge5
Switch(config-ge5)#gvrp enable
```

12.3.4 GVRP or MVRP Registration Mode

【Command】

```
(gvrp | mvrp) registration (fixed| forbidden | normal)
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

fixed: Fixed mode, no dynamic VLAN registration on the port, only static VLAN declaration messages are sent.

forbidden: Forbidden mode, does not allow dynamic VLAN to register on the port, simultaneously deletes all VLANs except VLAN 1 on the port, only sends VLAN 1 declaration message.

normal: normal mode, which allows dynamic VLANs to be registered on the port and simultaneously sends both static and dynamic VLAN declaration messages.

【Description】

(gvrp | mvrp) registration: this command is used for the GVRP port registration mode.

By default, the GVRP port registration mode is normal.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#gvrp enable
Switch(config)#interface ge5
Switch(config-ge5)#gvrp registration normal
```

12.3.5 GVRP or MVRP Timer

【Command】

(gvrp | mvrp) timer (join| leave| leaveall) <TIMER_VALUE>

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

leaveall | join | leave: represent leave All, join and leave three timers respectively. After each port starts GVRP/MVRP, it starts the LeaveAll timer at the same time, and the port will send the LeaveAll message circularly to make other ports re-register all their attribute information. GVRP/MVRP port can send each Join packet out twice to ensure the reliable transmission of messages, and the time interval between the two transmissions is controlled by the Join timer. GVRP/MVRP port can send each Join packet out twice to ensure the reliable transmission of messages, and the time interval between the two transmissions is controlled by the Join timer.

<TIMER_VALUE> : timer value, leave All defaults to 1000; The default value for join is 20; The default value for leave is 60. Unit: centiseconds.

【Description】

(gvrp | mvrp) timer: this command is used to configure leave All, join, and leave timers of the GARP ports.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#gvrp enable
Switch(config)#interface ge5
Switch(config-ge5)#gvrp timer join 10
```

12.3.6 Display Dynamic VLAN Information

【Command】

show vlan dynamic

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show vlan dynamic: this command is used to display dynamic VLAN information.

【Instance】

```
Switch> enable
Switch#show vlan dynamic
```

12.3.7 Display GVRP or MVRP Configuration Information

【Command】

show (gvrp | mvrp) configuration

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show gvrp | mvrp configuration: this command is used to display GVRP| MVRP configuration information.

【Instance】

```
Switch> enable
Switch#show gvrp configuration
```

12.3.8 Display GVRP or MVRP State Machine Information

【Command】

show (gvrp | mvrp) machine

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show gvrp | mvrp machine: this command is used to display GVRP| MVRP state machine information.

【Instance】

```
Switch> enable
Switch#show mvrp machine
```

12.3.9 Display GVRP or MVRP Message Statistics

【Command】

show (gvrp | mvrp) statistics [<IFNAME>]

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

Ifname：port name.

【Description】

show gvrp | mvrp statistics: this command is used to display GVRP| MVRP message statistics.

【Instance】

```
Switch> enable
Switch#show mvrp statistics
```

12.3.10 Display GVRP or MVRP Timer Information

【Command】

show (gvrp | mvrp) timer <IFNAME>

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

Ifname：port name.

【Description】

show gvrp | mvrp timer: this command is used to display timer information of GVRP| MVRP port .

【Instance】

```
Switch> enable
Switch#show mvrp time ge1
```

13 PIM-DM Configuration

13.1 Overview

PIM (Protocol Independent Multicast) is called Protocol Independent Multicast. By protocol agnostic, it means that the PIM does not need to maintain specialized unicast routing information. As a multicast routing solution, it directly uses the routing information of unicast routing table to perform RPF(Reverse Path Forwarding) check on multicast messages, and creates multicast routing table entries after passing the check, thus forwarding multicast messages.

Currently, Huawei switches support PIM-DM and PIM-SM.

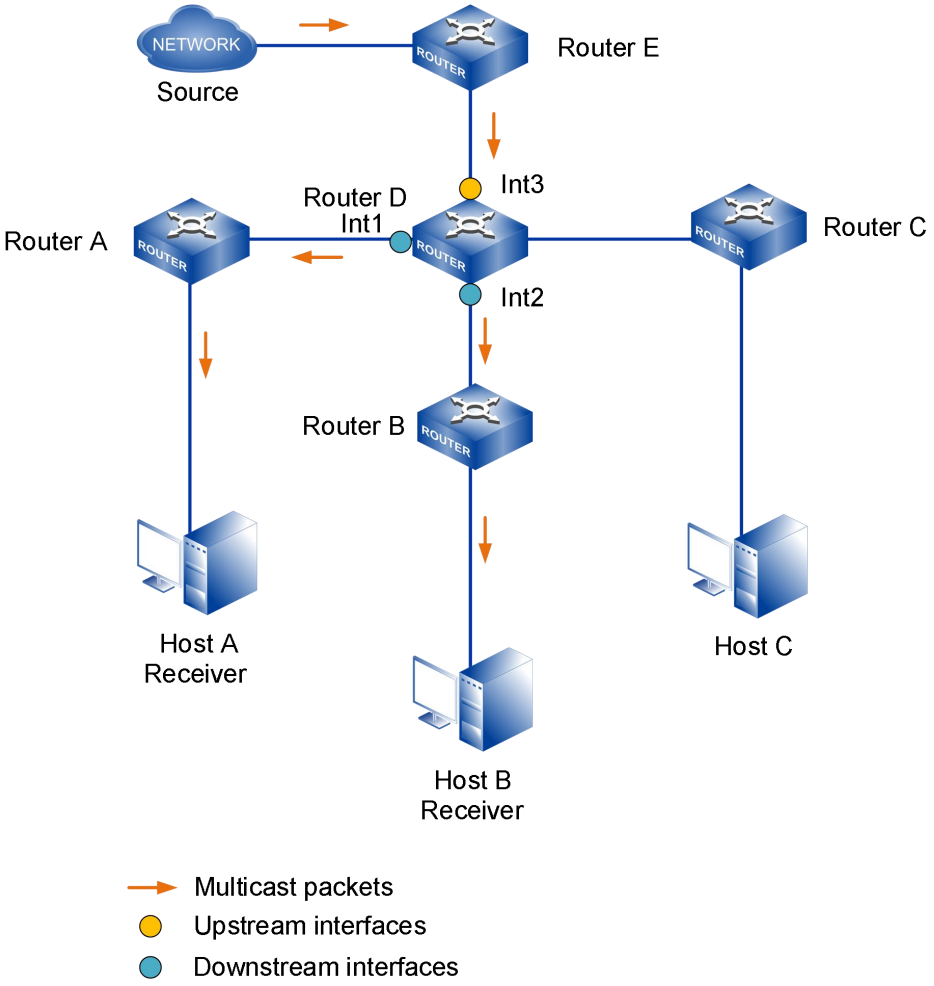
The multicast backbone (Mbone) was established in 1992 as a virtual backbone for IP multicast traffic to support network video and audio conferencing. Mbone uses Multicast Open shortest Path First (MOSPF) and the Distance Vector Multicast Routing Protocol (DVMRP).

Mbone promotes the use and development of multicast technology and routing multicast protocols. However, new problems arise: If multicast routes need to be created using multiple routing protocols, different protocols need to import routes of each other, which involves complex operations. Additionally, network devices need to maintain both unicast and multicast routing information. This creates a multicast routing protocol that has nothing to do with unicast - PIM. The PIM protocol focuses only on group membership and multicast source status-related information, while the selected path information is obtained directly from the unicast routing table. Instead, network devices only need to main group member and source information and obtain routing information from the unicast routing table. PIM is most widely used intra-domain multicast protocol.

A PIM network is comprised of multiple PIM routers. A large PIM network can be partitioned into multiple PIM domains to manage and control multicast packet forwarding.

13.2 Principle Description

13.2.1 Concepts



13.2.1.1 Multicast Distribution Tree

On a PIM network, a point-to-multipoint (P2MP) multicast forwarding path is established for each multicast group. The multicast forwarding path is also called a multicast distribution tree (MDT) because of its tree shape.

Two types of MDT are available:

- Shortest path tree (SPT) uses the multicast source as the root and multicast group members as leaves. SPT applies to both PIM-DM and PIM-SM networks.

The RouterE→RouterD→RouterA (RouterB) in the figure above is an SPT with root of Source and leaves of HostA and HostB.

- Rendezvous point tree (RPT) uses a rendezvous point (RP) as the root and multicast group members as leaves. RPT applies to PIM-SM networks.

13.2.1.2 PIM Router

Routers with PIM enabled on interfaces are called PIM routers. Roles of routers on an MDT include:

- Leaf router A router that connects to user hosts, which may not be multicast group members. RouterA, RouterB, and RouterC in above figure are leaf routers.
- First-hop router A router that directly connects to the multicast source on the multicast forwarding path and is responsible for forwarding multicast data from the multicast source. As RouterE in above figure.
- Last-hop router A router that directly connects to multicast group members (receivers) on the multicast forwarding path and is responsible for forwarding multicast data to these members. As RouterA and RouterB in above figure.
- Intermediate router A router located between the first-hop and last-hop routers. As RouterD in above figure.

13.2.1.3 PIM Routing Entry

PIM routing table entries are multicast routing table entries established by the PIM protocol. There are two routing table items in PIM network :(S, G) routing table items or (*, G) routing table items. (S, G) and (*, G), where S indicates a specific multicast source, G indicates a specific multicast group, and * indicates any multicast source.

- (S, G) Routing table entries are mainly used to establish SPT in the PIM network. It is suitable for PIM-DM network and PIM-SM network.
- (*, G) Routing table entries are mainly used to establish RPT in THE PIM network. Applicable to PIM-SM network.

A PIM router may have both (S, G) and (*, G) entries. When a PIM router receives a multicast packet with source address S and group address G and the packet passes the RPF check, the router forwards the packet according to the following rules:

- If a matching (S, G) entry exists, the router forwards the packet according to the (S, G) entry.
- If no matching (S, G) entry exists but a matching (*, G) entry exists, the router creates an (S, G) entry based on the (*, G) entry, and forwards the packet according to the (S, G) entry.

PIM routing entries contain the following information to guide multicast packet forwarding:

- Multicast source address
- Multicast group address
- Upstream interface: The interface on which multicast data is received on the local router, such as Int3 in the figure above.
- Downstream interface, which forwards multicast data, such as Int1 and Int2 in Figure above.

13.2.2 PIM-DM

13.2.2.1 Principles

PIM-DM forwards multicast packets in push mode and is used on small-scale networks with densely distributed multicast group members. PIM-DM assumes that each network segment has multicast group members. When a multicast source sends multicast packets, PIM-DM floods all PIM routers on the network with the multicast packets and prunes the branches that do not have multicast group members. Through periodic flooding and pruning, PIM-DM creates and maintains a unidirectional loop-free SPT that connects the multicast source and group members. If a new member joins a multicast group on the network segment connected to a leaf router on a pruned branch, the router can initiate the graft mechanism before starting new flooding and prune. The pruned branch turns into a forwarding branch.

PIM-DM uses neighbor discovery, flood, prune, graft, assert, and state refresh mechanisms. The flood, prune, and graft mechanisms are used to establish an SPT.

13.2.2.2 Neighbor Discovery

Each interface on the PIM router that enables the PIM protocol sends out Hello messages. The destination address of the group broadcast wrapping Hello message is 224.0.0.13 (representing all PIM routers in the same network segment), the source address is the IP address of the interface, and the TTL value is 1.

Hello messages are used to discover PIM neighbors, adjust PIM protocol parameters, and maintain neighbor relationships.

- Discovering PIM neighbors

  All PIM routers on the same network segment must receive Hello messages. By exchanging Hello messages, directly connected PIM routers learn neighbor information and establish neighbor relationships.

  A PIM router can receive other PIM messages to create multicast routing entries only after it establishes neighbor relationships with other PIM routers.

- Adjusting PIM protocol parameters

  Hello message carries a number of PIM protocol message parameters, which are mainly used for the control of PIM protocol messages between neighbors of PIM. Details are as follows:

  - DR_Priority: indicates the priority used for designated router (DR) election among interfaces. The interface with the highest priority becomes the DR. The DR_Priority parameter is only used in DR election on PIM-SM networks.
  - Holdtime: represents the timeout period for keeping the neighbor in the reachable state. If no Hello message from the PIM neighbor is received within the timeout period, the router considers the neighbor unreachable.
  - LAN_Delay: indicates the delay in transmitting Prune messages on a shared network segment.
  - Neighbor-Tracking: indicates the neighbor tracking function.
  - Override-Interval: indicates the interval for overriding the prune mechanism.
- Maintaining neighbor relationships

  Between PIM routers Hello messages are sent periodically. If the Holdtime timeout has not yet received a new Hello message from the PIM neighbor, the PIM router considers the neighbor unreachable and clears it from the neighbor list.

  Changes of PIM neighbors lead to multicast topology changes on the network. If an upstream or downstream neighbor in the MDT is unreachable, multicast routes re-converge and the MDT is re-established.

13.2.2.3 Flooding

On a PIM-DM network, multicast packets from a multicast source are flooded throughout the entire network. When a PIM router receives a multicast packet, it uses unicast routing information to perform an RPF check. If the packet passes the RPF check, the router creates an (S, G) entry, in which the downstream interface list contains all the interfaces connected to PIM neighbors, except the interface connected to the upstream router. The router then forwards subsequent multicast packets through each downstream interface.

When the multicast packets reach a leaf router, the leaf router processes the packets as follows:

- If the network segment connected to the leaf router has group members, the leaf router adds its interface that is connected to the network segment to the downstream interface list of the (S, G) entry, and forwards subsequent multicast

packets to the group members.

- If there is no group member in the network segment of the user connected to the leaf router, and there is no need to forward the group broadcast to its downstream PIM neighbor, a Prune action is performed.

Multicast packets are sometimes flooded to a shared network segment with multiple PIM routers. If the packets pass the RPF check on these PIM routers, multiple copies of multicast packets are forwarded to this network segment. These PIM routers will need to initiate the assert mechanism.

13.2.2.4 Pruning

When a PIM router receives a multicast packet, it performs an RPF check on the packet. If the packet passes the RPF check but the downstream network segment does not need to receive the multicast packet, the PIM router sends a Prune message to the upstream router. After receiving the Prune message, the upstream router deletes the downstream interface from the downstream interface list of the created (S, G) entry and no longer forwards multicast packets to the downstream interface. A leaf router initiates the prune mechanism, and the Prune message is sent upstream by hop along the MDT to prune the network segment with no group members.

The PIM router starts a prune timer for the pruned downstream interface. The interface resumes forwarding multicast packets after the timer expires. Subsequently, multicast packets are flooded throughout the entire network and new group members can receive multicast packets. If a leaf router connected to a network segment that has no group members receives the flooded multicast packets, the leaf router initiates the prune mechanism. PIM-DM updates the SPT periodically through the process of periodic flooding and pruning.

When the downstream interface is pruned:

- Grafting is performed if the downstream leaf router has a group that joins and wishes to resume group broadcast forwarding before the next diffuse-pruning.
- When no member joins a multicast group on the network segment connected to the leaf router and the downstream interface still needs to be suppressed, the leaf router initiates the state refresh mechanism.

13.2.2.5 Graft

PIM-DM uses the graft mechanism to enable new group members on a pruned network segment to rapidly obtain multicast data. IGMP helps a leaf router learn whether new group members have joined a multicast group on the connected network

segment. If a leaf router learns that new group members have joined multicast group G, the leaf router sends a Graft message to the upstream router. The message requests the upstream router to resume multicast packet forwarding on the downstream interface and to add the downstream interface to the downstream interface list of the (S, G) entry.

A graft process is initiated by a leaf router and ends at the upstream router that receives the multicast packets.

13.2.2.6 State Refresh

To prevent a pruned interface from resuming multicast packet forwarding after the prune timer expires, the first-hop router nearest to the multicast source periodically sends a State-Refresh message throughout the entire PIM-DM network. PIM routers receiving the State-Refresh message refresh the prune timer state. If no group member joins a multicast group on the network segment connected to a leaf router in a pruned branch, the upstream interface connected to this router remains suppressed.

13.2.2.7 Assert

When multicast packets pass the RPF check on multiple PIM routers connecting to a network segment, the assert mechanism is required to ensure that only one PIM router forwards the multicast packets to the network segment. When a PIM router receives a multicast packet that is the same as the multicast packet it sends to other neighbors, the PIM router sends an Assert message with the destination address 224.0.0.13 to all other PIM routers on the same network segment. The PIM routers carry out assert election by comparing their parameters to those carried in the received Assert message. The election rules are as follows:

1 If these routers have different unicast routing priorities, the router with the highest unicast routing priority wins.

2 If these routers have the same unicast routing priority, the router with the smallest route cost to the multicast source wins.

3 If these routers have the same unicast routing priority and the same route cost to the multicast source, the router with the highest IP address for the downstream interface wins.

A PIM router performs the following operations based on assert election results:

- If a router wins the assert election, its downstream interface becomes the assert winner and is responsible for forwarding multicast packets to the shared network segment.

- If a router loses the assert election, its downstream interface becomes an assert loser, which is deleted from the downstream interface list of the (S, G) entry and no longer forward multicast packets to the shared network segment.

After an assert election is complete, only one downstream interface exists on the shared network segment and it transmits only one copy of multicast packets. All assert losers periodically resume multicast packet forwarding, causing an assert election each time.

13.3 Configure PIM-DM

13.3.1 Enable PIM-DM

【Command】

ip pim dense-mode

no ip pim dense-mode

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip pim dense-mode: this command is used to enable PIM-DM.

no ip pim dense-mode: this command is used to disable the PIM-DM.

By default, PIM-DM is disabled.

Notice

- This command takes effect only if IP multicast routing is globally enabled.
- For configuration, see the command ip multicast - routing.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim dense-mode
```

13.3.2 PIM-DM DR Priority

【Command】

ip pim dense-mode dr-priority PRIORITY

no ip pim dense-mode dr-priority

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

PRIORITY: specifies the priority of campaign DR, with the value range from 0 to 4294967294. The higher the value, the higher the priority.

【Description】

ip pim dense- mode dr-priority: this command is used to configure the priority of the campaign DR on the interface.

no ip pim dense-mode dr-priority: this command is used to restore to the default value.

By default, campaign DR has priority of 1.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim dense-mode dr-priority 10
```

13.3.3 The PIM - DM GenID Information

【Command】

ip pim dense-mode exclude-genid

no ip pim dense-mode exclude-genid

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip pim dense-mode exclude-genid: this command is used to configure the interface to send hello messages without GenID information.

no ip pim dense-mode exclude-genid: this command is used to restore to the default value.

By default, the hello message from the interface carries GenID information.

GenID is a random value at the initial creation of an interface to identify unique interface information. With this information, users can detect whether the neighbor device has been restarted.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim dense-mode exclude-genid
```

13.3.4 PIM-DM Neighbor Reachable State Time

【Command】

ip pim dense-mode hello-holdtime TIME

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

TIME: specifies the time to maintain the reachable state of the PIM neighbor. The value ranges from 1 to 65535 in seconds. If specified as 65535 seconds, the PIM neighbor is always reachable.

【Description】

ip pim dense-mode hello-holdtime:this command is used to configure the time on the interface to keep the reachable state of the PIM neighbor.

no ip pim dense-mode hello-holdtime: this command is used to restore to the default value.

By default, the PIM neighbor is kept in the reachable state for 105 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim dense-mode hello-holdtime 150
```

13.3.5 Time Interval of PIM-DM Hello Message

【Command】

```
ip pim dense-mode hello-interval INTERVAL
no ip pim dense-mode hello-interval
```

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

INTERVAL: Specifies the time interval between sending Hello message and the value ranges from 1 to 65535 in seconds.

【Description】

ip pim dense-mode hello-interval: this command is used to configure the time interval for sending Hello messages on the interface.

no ip pim dense-mode hello-interval: this command is used to restore to the default value.

By default, Hello messages are sent at an interval of 30 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim dense-mode hello-interval 60
```

13.3.6 PIM-DM Neighbor Filter

【Command】

```
ip pim dense-mode neighbor-filter <ACL-NUMBER >
```

no ip pim dense-mode neighbor-filter

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

ACL- NUMBER: standard ACL number, ranging from 1-99 or 1300-1999.

【Description】

ip pim dense-mode neighbor-filter: this command is used to configure an illegal neighbor source address range. The action for the ACL must be deny, and if the action is permit, a mismatch is considered a legitimate neighbor.

no ip pim dense-mode neighbor-filter: this command is used to cancel configuration of an illegal neighbor source address range.

By default, there are no restrictions on the neighbor source addresses that an interface can learn from.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 20 deny 192.168.1.0 0.0.0.255
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim dense-mode neighbor-filter 20
```

13.3.7 IPM-DM Status Refresh Message Enablement

【Command】

ip pim dense-mode state-refresh

no ip pim dense-mode state-refresh

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip pim dense-mode state-refresh: this command is used for global configuration to enable sending a status refresh message.

no ip pim dense-mode state-refresh: this command is used to cancel the restore configuration.

By default, the global enable sends a state refresh message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#no ip pim dense-mode state-refresh
```

13.3.8 The Time Interval During Which PIM-DM Sends a Status Refresh Message

【Command】

ip pim dense-mode state-refresh origination-interval < VALUE >
no ip pim dense-mode state-refresh origination-interval

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

VALUE: The time interval between sending a status refresh, with a value range of 1-100 in seconds.

【Description】

ip pim dense-mode state-refresh origination-interval: this command is used to globally set the time interval for sending a status refresh message.

no ip pim dense-mode state-refresh origination-interval: this command is used to restore to the default.

By default, a state refresh is sent at an interval of 60 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim dense-mode state-refresh
origination-interval 30
```

13.3.9 The Time Interval Between PIM-DM Receiving a Status Refresh Message

【Command】

ip pim dense-mode state-refresh rate limit-interval < VALUE >

no ip pim dense-mode state**-refresh limit-interval**

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

VALUE: The time interval between receiving the state refresh message, the value range is 1 ~ 100, and the unit is seconds/piece.

【Description】

ip pim dense-mode state-refresh rate limit-interval: this command is used to globally set the time interval for receiving state refresh messages.

no ip pim dense-mode state-refresh rate limit-interval: this command is used to restore to the default.

By default, the time interval for receiving the status refresh message is 30 seconds per message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim dense-mode state-refresh rate limit-interval 60
```

13.3.10 Display PIM-DM Detail Information

【Command】

show ip pim dense-mode detail

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

This command is used to view PIM details for the device.

【Instance】

```
Switch#show ip pim dense-mode detail

PIM Interface information:


PIM Local-Membership information:

PIM Neighbor information:


IP Multicast Route Table:
```

13.3.11 Display PIM-DM Interface Information

【Command】

show ip pim dense-mode interfce [detail]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

detail: specifies to view the details.

【Description】

The command is used to view PIM information on the interface.

【Instance】

```
Switch#show ip pim dense-mode interface
Address         Interface VIFindex  Ver/    Nbr    DR    DR
                                    Mode   Count  Prior
10.3.101.252        vlanif10    0            v2/S     1          1
10.3.101.253
11.1.1.2         vlanif11  2       v2/S   1      1     11.1.1.3
172.24.0.2       vlanif50  3       v2/S   1      1     172.24.0.3
```

```
Switch#show ip pim dense-mode interface detail
vlanif10 (vif 0):
  Address 10.3.101.252, DR 10.3.101.253
  Hello period 30 seconds, Next Hello in 2 seconds
  Triggered Hello period 5 seconds
  Neighbors:
   10.3.101.253

vlanif11 (vif 2):
  Address 11.1.1.2, DR 11.1.1.3
  Hello period 30 seconds, Next Hello in 2 seconds
  Triggered Hello period 5 seconds
  Secondary addresses:
   11.1.1.1
  Neighbors:
   11.1.1.3

vlanif50 (vif 3):
  Address 172.24.0.2, DR 172.24.0.3
  Hello period 30 seconds, Next Hello in 3 seconds
  Triggered Hello period 5 seconds
  Neighbors:
   172.24.0.3
```

13.3.12 Displays the Local Multicast Group Members of the PIM-DM

【Command】

show ip pim dense-mode local-members [<IFNAME>]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: Specifies the vlanif interface.

【Description】

View the local multicast group members of the PIM-DM for the current device.

【Instance】

```
Switch#show ip pim dense-mode local-members vlanif10
PIM Local membership information
vlanif10:
  (*, 225.2.2.2) : Include
```

13.3.13 Displays PIM-DM Multicast Routing Table Entries

【Command】

show ip pim dense-mode mroute [detail]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

detail: specifies to view the details.

【Description】

View the current PIM-DM multicast routing table entry.

【Instance】

None

13.3.14 Display PIM-DM Neighbor Information

【Command】

show ip pim dense-mode neighbor [detail]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

detail: specifies to view the details.

【Description】

View the PIM-DM neighbor running status of the current learning.

【Instance】

```
Switch#show ip pim dense-mode neighbor
Neighbor        Interface         Uptime/Expires    Ver   DR
Address                                             Priority/Mode
10.3.101.253      vlanif10          01:53:56/00:01:17  v2    1 /
DR
11.1.1.3         vlanif11          01:53:57/00:01:18  v2    1 / DR
172.24.0.3       vlanif50          01:53:56/00:01:17  v2    1 / DR
```

13.3.15 Displays PIM-DM Next Hop Information

【Command】

show ip pim dense-mode nexthop

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

View the next hop information used by the PIM-DM module in the current device.

【Instance】

```
Switch#show ip pim dense-mode nexthop
Flags: N = New, R = RP, S = Source, U = Unreachable
Destination  Type  Nexthop   Nexthop      Nexthop  Nexthop Metric
Pref  Refcnt
               Num     Addr         Ifindex    Name
_____________________________________________________________________
_____________
10.3.101.111  ..S.  1         0.0.0.0         4                0
0     2
```

14 PIM-SM Configuration

14.1 Overview

PIM (Protocol Independent Multicast) is called Protocol Independent Multicast. By protocol agnostic, it means that the PIM does not need to maintain specialized unicast routing information. As a multicast routing solution, it directly uses the routing information of unicast routing table to perform RPF(Reverse Path Forwarding) check on multicast messages, and creates multicast routing table entries after passing the check, thus forwarding multicast messages.

Currently, Huawei switches support PIM-DM and PIM-SM.

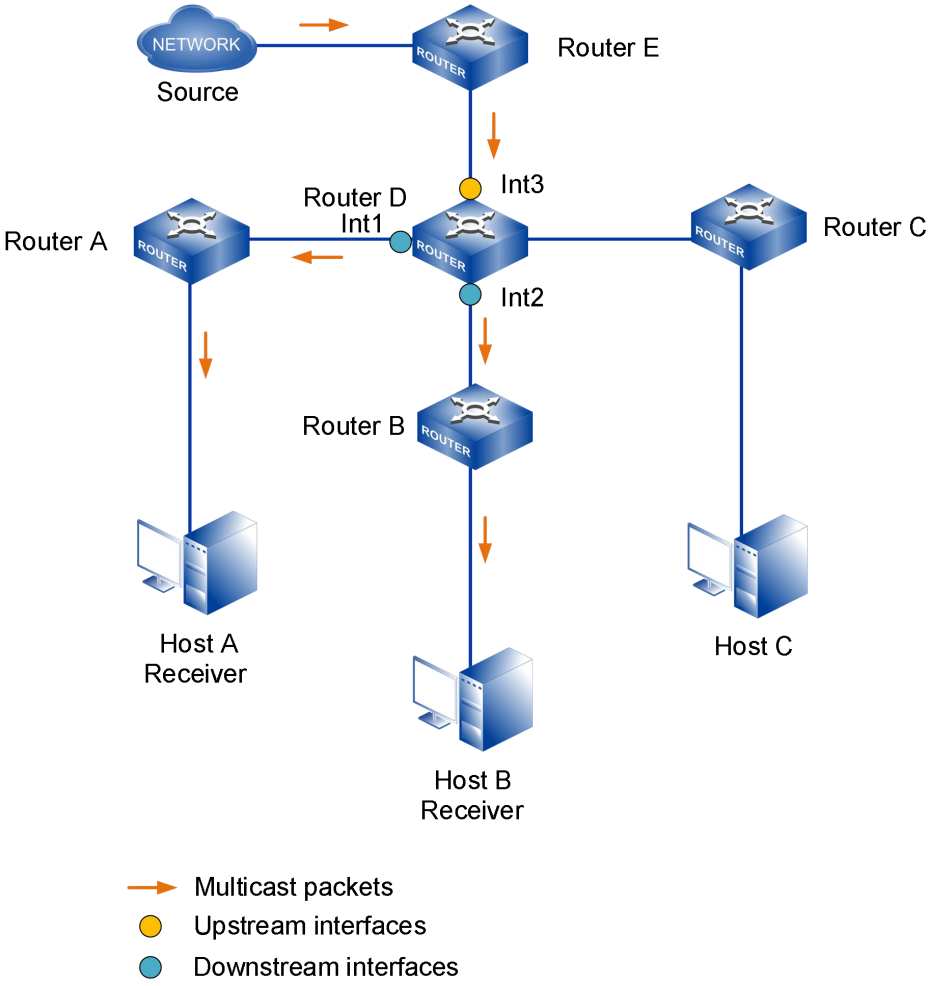
The multicast backbone (Mbone) was established in 1992 as a virtual backbone for IP multicast traffic to support network video and audio conferencing. Mbone uses Multicast Open shortest Path First (MOSPF) and the Distance Vector Multicast Routing Protocol (DVMRP).

Mbone promotes the use and development of multicast technology and routing multicast protocols. However, new problems arise: If multicast routes need to be created using multiple routing protocols, different protocols need to import routes of each other, which involves complex operations. Additionally, network devices need to maintain both unicast and multicast routing information. This creates a multicast routing protocol that has nothing to do with unicast - PIM. The PIM protocol focuses only on group membership and multicast source status-related information, while the selected path information is obtained directly from the unicast routing table. Instead, network devices only need to main group member and source information and obtain routing information from the unicast routing table. PIM is most widely used intra-domain multicast protocol.

A PIM network is comprised of multiple PIM routers. A large PIM network can be partitioned into multiple PIM domains to manage and control multicast packet forwarding.

14.2 Principle Description

14.2.1 Concepts



14.2.1.1 Multicast Distribution Tree

On a PIM network, a point-to-multipoint (P2MP) multicast forwarding path is established for each multicast group. The multicast forwarding path is also called a multicast distribution tree (MDT) because of its tree shape.

Two types of MDT are available:

- Shortest path tree (SPT) uses the multicast source as the root and multicast group members as leaves. SPT applies to both PIM-DM and PIM-SM networks.

The RouterE→RouterD→RouterA (RouterB) in the figure above is an SPT with root of Source and leaves of HostA and HostB.

- Rendezvous point tree (RPT) uses a rendezvous point (RP) as the root and multicast group members as leaves. RPT applies to PIM-SM networks.

14.2.1.2 PIM Router

Routers with PIM enabled on interfaces are called PIM routers. Roles of routers on an MDT include:

- Leaf router A router that connects to user hosts, which may not be multicast group members. RouterA, RouterB, and RouterC in above figure are leaf routers.
- First-hop router A router that directly connects to the multicast source on the multicast forwarding path and is responsible for forwarding multicast data from the multicast source. As RouterE in above figure.
- Last-hop router A router that directly connects to multicast group members (receivers) on the multicast forwarding path and is responsible for forwarding multicast data to these members. As RouterA and RouterB in above figure.
- Intermediate router A router located between the first-hop and last-hop routers. As RouterD in above figure.

14.2.1.3 PIM Routing Entry

PIM routing table entries are multicast routing table entries established by the PIM protocol. There are two routing table items in PIM network :(S, G) routing table items or (*, G) routing table items. (S, G) and (*, G), where S indicates a specific multicast source, G indicates a specific multicast group, and * indicates any multicast source.

- (S, G) Routing table entries are mainly used to establish SPT in the PIM network. It is suitable for PIM-DM network and PIM-SM network.
- (*, G) Routing table entries are mainly used to establish RPT in THE PIM network. Applicable to PIM-SM network.

A PIM router may have both (S, G) and (*, G) entries. When a PIM router receives a multicast packet with source address S and group address G and the packet passes the RPF check, the router forwards the packet according to the following rules:

- If a matching (S, G) entry exists, the router forwards the packet according to the (S, G) entry.
- If no matching (S, G) entry exists but a matching (*, G) entry exists, the router creates an (S, G) entry based on the (*, G) entry, and forwards the packet according to the (S, G) entry.

PIM routing entries contain the following information to guide multicast packet forwarding:

- Multicast source address
- Multicast group address
- Upstream interface: The interface on which multicast data is received on the local router, such as Int3 in the figure above.
- Downstream interface, which forwards multicast data, such as Int1 and Int2 in Figure above.

14.2.2 PIM-SM (ASM Model)

14.2.2.1 Principles

In Any-Source Multicast (ASM) implementation, PIM-Sparse Mode (PIM-SM) forwards multicast packets in pull mode and is for use on large-scale networks with sparsely distributed group members. Devices on the PIM-SM network work as follows:

- A Rendezvous Point (RP), an important PIM router, is available to provide services for group members or multicast sources that appear at any time. All PIM routers on the network know the position of the RP.
- When a user host joins a multicast group G using IGMP, the last-hop router sends a Join message to the RP. A (*, G) entry is created hop by hop, and an RPT with the RP as the root is generated.
- When a multicast source sends the first multicast packet to a multicast group G, the first-hop router encapsulates the multicast data in a Register message and sends the Register message to the RP in unicast mode. The RP then creates an (S, G) entry and registers multicast source information.

In the ASM model, PIM-SM uses the neighbor discovery, DR election, RP discovery, RPT setup, multicast source registration, SPT switchover, and assertion mechanisms. A Bootstrap router (BSR) can also be configured to implement fine-grained management in a single PIM-SM domain.

14.2.2.2 Neighbor Discovery

The neighbor discovery mechanism is the same as in PIM-DM.

Each interface on the PIM router that enables the PIM protocol sends out Hello messages. The destination address of the group broadcast wrapping Hello message is 224.0.0.13 (representing all PIM routers in the same network segment), the source address is the IP address of the interface, and the TTL value is 1.

Hello messages are used to discover PIM neighbors, adjust PIM protocol parameters, and maintain neighbor relationships.

- Discovering PIM neighbors

All PIM routers on the same network segment must receive Hello messages. By exchanging Hello messages, directly connected PIM routers learn neighbor information and establish neighbor relationships.

A PIM router can receive other PIM messages to create multicast routing entries only after it establishes neighbor relationships with other PIM routers.

- Adjusting PIM protocol parameters

  Hello message carries a number of PIM protocol message parameters, which are mainly used for the control of PIM protocol messages between neighbors of PIM. Details are as follows:

  - DR_Priority: indicates the priority used for designated router (DR) election among interfaces. The interface with the highest priority becomes the DR. The DR_Priority parameter is only used in DR election on PIM-SM networks.
  - Holdtime: represents the timeout period for keeping the neighbor in the reachable state. If no Hello message from the PIM neighbor is received within the timeout period, the router considers the neighbor unreachable.
  - LAN_Delay: indicates the delay in transmitting Prune messages on a shared network segment.
  - Neighbor-Tracking: indicates the neighbor tracking function.
  - Override-Interval: indicates the interval for overriding the prune mechanism.

- Maintaining neighbor relationships

  Between PIM routers Hello messages are sent periodically. If the Holdtime timeout has not yet received a new Hello message from the PIM neighbor, the PIM router considers the neighbor unreachable and clears it from the neighbor list.

  Changes of PIM neighbors lead to multicast topology changes on the network. If an upstream or downstream neighbor in the MDT is unreachable, multicast routes re-converge and the MDT is re-established.

14.2.2.3 DR Election

The network segment where a multicast source or group member resides is usually connected to multiple PIM routers. These PIM routers exchange Hello messages to set up PIM neighbor relationships. The Hello messages carry the DR priority and the interface address of the network segment. Each PIM router compares its own information with the information carried in messages sent by its neighbors. A DR is

then elected for multicast data forwarding on the source side or the receiver side. The election rules are as follows:

- If all PIM routers on the network segment allow Hello messages to carry DR priorities, the PIM router with the highest DR priority is elected as the DR.
- If multiple PIM routers have the same DR priority, or if at least one PIM router does not allow Hello messages to carry the DR priority, the PIM router with the largest IP address is elected as the DR.

If an existing DR becomes faulty, PIM neighbor relationships time out, and a new DR election is triggered among PIM neighbors.

14.2.2.4 RP Discovery

An RP is responsible for processing Register messages from the multicast source and Join messages from group members. All PIM routers on the network know the location of the RP.

An RP can serve multiple multicast groups simultaneously, but each multicast group can be associated with only one RP. RPs can be configured statically or elected dynamically:

- Static RP All PIM routers on the network are manually configured with the same RP address.
- Dynamic RP Several PIM routers in a PIM domain are configured as candidate RPs (C-RPs) and an RP is elected from them. One or more PIM routers are configured as candidate BSRs (C-BSRs), and a BSR from them to collect and advertise C-RP information.

  During the C-BSR campaign, at the beginning, every C-BSR considered himself as BSR and sent Bootstrap messages to the whole network. Bootstrap messages carry C-BSR address and C-BSR priority. Each PIM router compares the Bootstrap messages it receives from C-BSRs. The election rules are as follows:

  - If the C-BSRs have different priorities, the C-BSR with the highest priority (largest priority value) is elected as the BSR.
  - If the C-BSRs have the same priority, the C-BSR with the largest IP address is elected as the BSR.

14.2.2.5 RPT Setup

PIM-SM RPT is a multicast distribution tree with RP as root and PIM router with group membership as leaf. When group members appear in the network (user hosts Join a multicast group G through IGMP), group member DR sends Join message to RP, and

creates (*, G) table items step by step on the path to RP to generate an RPT with RP as the root.

In the process of RPT construction, when the PIM router sends Join message, it will conduct RPF check: search the unicast route to RP. The egress interface of the unicast route is the upstream interface and the next hop is the RPF neighbor. Join message is sent hopped from group member DR to RP.

14.2.2.6 Multicast Source Registration

In the PIM-SM network, any new multicast source must first be "registered" in the RP before it can be transmitted to the group members. The registration and forwarding process is as follows:

1 The multicast source sends a multicast packet to the source DR.

2 After receiving the multicast packet, the source DR encapsulates the multicast packet into a Register message and sends the Register message to the RP.

3 After receiving the Register message, the RP decapsulates it, creates an (S, G) entry, and sends multicast data to group members along the RPT.

14.2.2.7 SPT Switchover

A multicast group on a PIM-SM network can be associated with only one RP and sets up only one RPT. On the condition of no SPT switching, all multicast packets destined for a multicast group are encapsulated in Register messages and sent to the RP. The RP then decapsulates the packets and forwards them along the RPT to multicast group members. RP is a necessary transit station for all multicast messages. when the multicast message rate gradually increases, it will create a huge burden on RP. PIM-SM allows the RP or the receiver DR to trigger an SPT switchover to solve the problem.

- SPT switchover triggered by the RP

  After receiving a Register message from the source DR, the RP forwards multicast packets **which is encapsulated in the Register message** along the RPT to group members, and send Join message to the source DR hop by hop The RP also sends a Join message to the source DR to set up an SPT from the RP to the source.

  After the SPT is set up, the source DR forwards multicast packets directly to the RP. After the switchover, the source DR and RP no longer need to encapsulate or decapsulate packets.

- SPT switchover triggered by the receiver DR
  When the receiver DR finds that the forwarding rate is higher than the configured threshold, it triggers an SPT switchover. The SPT switchover is as follows:
  - The receiver DR sends a Join message to the source DR hop by hop, creates an (S, G) entry hop by hop, and then sets up an SPT from the source DR to the receiver DR.
  - After the SPT is set up, the receiver DR sends Prune messages along the RPT to the RP, which then deletes the downstream interface connected to the receiver DR from the (*, G) entry. After the prune action is complete, the RP no longer forward multicast packets along the RPT.
  - Because the SPT does not pass through the RP, the RP continues to send Prune messages along the RPT to the source DR, which then deletes the downstream interface connected to the RP from the (S, G) entry. After the prune action is complete, the source DR no longer forward multicast packets along the SPT to the RP.

14.2.2.8 Assert

The assert mechanism is the same as in PIM-DM.

When multicast packets pass the RPF check on multiple PIM routers connecting to a network segment, the assert mechanism is required to ensure that only one PIM router forwards the multicast packets to the network segment. When a PIM router receives a multicast packet that is the same as the multicast packet it sends to other neighbors, the PIM router sends an Assert message with the destination address 224.0.0.13 to all other PIM routers on the same network segment. The PIM routers carry out assert election by comparing their parameters to those carried in the received Assert message. The election rules are as follows:

1 If these routers have different unicast routing priorities, the router with the highest unicast routing priority wins.

2 If these routers have the same unicast routing priority, the router with the smallest route cost to the multicast source wins.

3 If these routers have the same unicast routing priority and the same route cost to the multicast source, the router with the highest IP address for the downstream interface wins.

A PIM router performs the following operations based on assert election results:

- If a router wins the assert election, its downstream interface becomes the assert winner and is responsible for forwarding multicast packets to the shared network segment.

- If a router loses the assert election, its downstream interface becomes an assert loser, which is deleted from the downstream interface list of the (S, G) entry and no longer forward multicast packets to the shared network segment.

After an assert election is complete, only one downstream interface exists on the shared network segment and it transmits only one copy of multicast packets. All assert losers periodically resume multicast packet forwarding, causing an assert election each time.

14.2.2.9 BSR Administrative Domain

To provide fine-grained network management, a PIM-SM network has both a global domain and multiple BSR administrative domains. This reduces the workload on individual BSRs and allows for the use of private group addresses to provide special services to users in specific domains.

Each BSR administrative domain maintains only one BSR, which serves multicast groups within a specific range. The global domain maintains a BSR that serves multicast groups not served by an administrative domain.

This section describes the relationship between BSR administrative domains and the global domain in terms of domain space, group address ranges, and multicast functions.

- Domain space

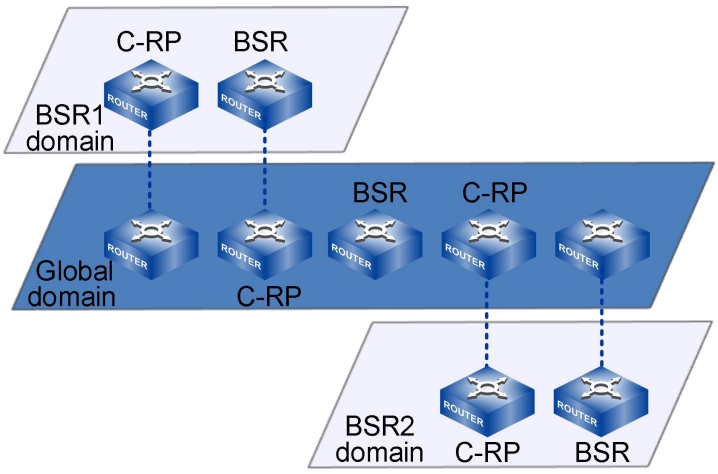
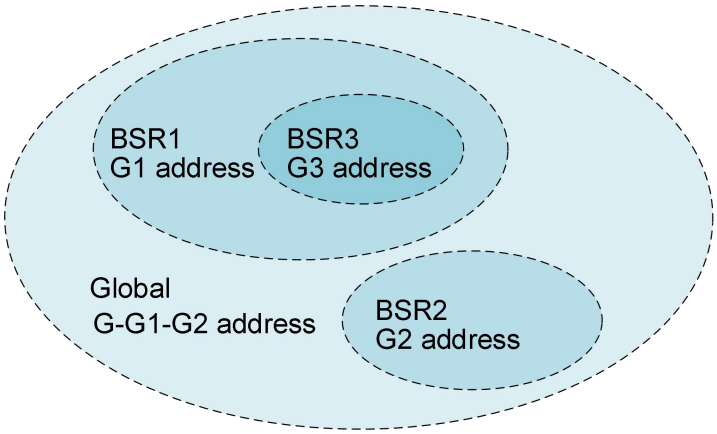


Figure above shows the relationship between BSR administrative domains and the global domain in terms of domain space. BSR administrative domains for the same group address contain different PIM routers. A PIM router belongs to only one BSR administrative domain. BSR administrative domains are independent of and isolated from each other. Each BSR administrative domain manages the

multicast groups within a specific group address range. Multicast packets within this range can be transmitted only within this administrative domain and cannot cross its border.

The global domain contains all PIM routers on the PIM-SM network. A multicast packet that does not belong to any BSR administrative domain can be transmitted throughout the entire PIM network.

- Group address range



Each BSR administrative domain serves multicast groups within a specific range. The group address ranges served by different BSR administrative domains can overlap. In Figure 7, the group address range of BSR1 overlaps that of BSR3. As shown in the figure above, the group address ranges corresponding to the BSR1 and BSR3 domains overlap.

Multicast groups that do not belong to any BSR management domain are all within the service scope of the Global domain. As shown in the figure above, the Global domain group address range is G-G1-G2 in addition to G1 and G2.

- Multicast function

As shown in the geographic space image, the Global domain and each BSR management domain contain C-RP and BSR devices for their own domain, which are only valid within their domain when they perform the corresponding functions. Each domain holds independent BSR and RP elections.

Each BSR administrative domain has a border. Multicast messages from this domain, such as C-RP Advertisement messages or BSR BootStrap messages, can be transmitted only within the domain from which they originate. Multicast messages from the global domain can be transmitted throughout the entire global domain and traverse any BSR administrative domain.

14.2.3 PIM-SM (SSM Model)

14.2.3.1 Principles

The SSM model uses IGMPv3/MLDv2 and PIM-SM technology. There is no need to maintain an RP, set up an RPT, or register a multicast source. An SPT can be built directly between the source and group members.

In the SSM model, user hosts know the locations of multicast sources before requesting multicast services. When user hosts join multicast groups, they can specify the sources from which they want to receive data. After receiving requests from user hosts, the receiver DR sends a Join message to the source DR. The Join message is then transmitted upstream hop by hop to set up an SPT between the source and group members.

In the SSM model, PIM-SM uses the neighbor discovery, DR election, and SPT setup mechanisms.

14.2.3.2 Neighbor Discovery

The neighbor discovery mechanism is the same as in PIM-DM.

Each interface on the PIM router that enables the PIM protocol sends out Hello messages. The destination address of the group broadcast wrapping Hello message is 224.0.0.13 (representing all PIM routers in the same network segment), the source address is the IP address of the interface, and the TTL value is 1.

Hello messages are used to discover PIM neighbors, adjust PIM protocol parameters, and maintain neighbor relationships.

- Discovering PIM neighbors

  All PIM routers on the same network segment must receive Hello messages. By exchanging Hello messages, directly connected PIM routers learn neighbor information and establish neighbor relationships.

  A PIM router can receive other PIM messages to create multicast routing entries only after it establishes neighbor relationships with other PIM routers.
- Adjusting PIM protocol parameters

  Hello message carries a number of PIM protocol message parameters, which are mainly used for the control of PIM protocol messages between neighbors of PIM. Details are as follows:
  - DR_Priority: indicates the priority used for designated router (DR) election among interfaces. The interface with the highest priority becomes the DR.

The DR_Priority parameter is only used in DR election on PIM-SM networks.

- Holdtime: represents the timeout period for keeping the neighbor in the reachable state. If no Hello message from the PIM neighbor is received within the timeout period, the router considers the neighbor unreachable.
- LAN_Delay: indicates the delay in transmitting Prune messages on a shared network segment.
- Neighbor-Tracking: indicates the neighbor tracking function.
- Override-Interval: indicates the interval for overriding the prune mechanism.

- Maintaining neighbor relationships

  Between PIM routers Hello messages are sent periodically. If the Holdtime timeout has not yet received a new Hello message from the PIM neighbor, the PIM router considers the neighbor unreachable and clears it from the neighbor list.

  Changes of PIM neighbors lead to multicast topology changes on the network. If an upstream or downstream neighbor in the MDT is unreachable, multicast routes re-converge and the MDT is re-established.

14.2.3.3 DR Election

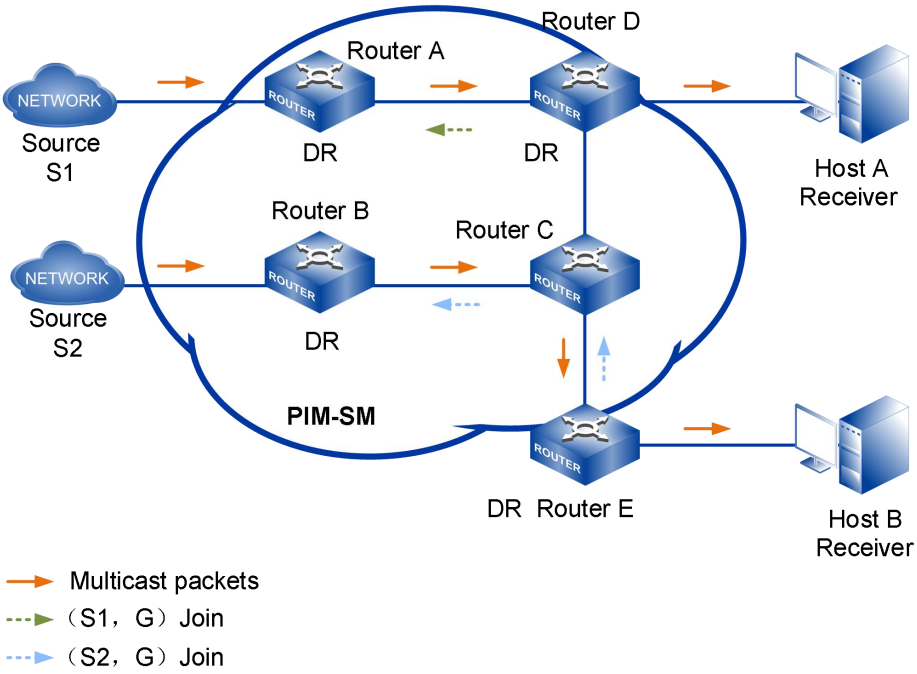
The DR election mechanism is the same as in the PIM-SM (ASM model).

The network segment where a multicast source or group member resides is usually connected to multiple PIM routers. These PIM routers exchange Hello messages to set up PIM neighbor relationships. The Hello messages carry the DR priority and the interface address of the network segment. Each PIM router compares its own information with the information carried in messages sent by its neighbors. A DR is then elected for multicast data forwarding on the source side or the receiver side. The election rules are as follows:

- If all PIM routers on the network segment allow Hello messages to carry DR priorities, the PIM router with the highest DR priority is elected as the DR.
- If multiple PIM routers have the same DR priority, or if at least one PIM router does not allow Hello messages to carry the DR priority, the PIM router with the largest IP address is elected as the DR.

If an existing DR becomes faulty, PIM neighbor relationships time out, and a new DR election is triggered among PIM neighbors.

14.2.3.4 SPT Setup



As shown in the figure above, the establishment process of SPT is as follows:

1 Using IGMPv3/MLDv2, RouterD and RouterE learn that packets from user hosts have the same multicast group address but are requesting multicast data from different source addresses. They send Join messages to sources hop by hop.

2 PIM routers create (S1, G) and (S2, G) entries based on the Join messages and set up SPTs from S1 to HostA and from S2 to HostB.

3 After SPTs are set up, the sources forward multicast packets along the SPTs to group members.

14.2.3.5 Comparison with the ASM Model

The biggest difference between the SSM model and the ASM model is whether a multicast source is specified, as shown in the table below.

Protocol	Model classification	Applicable scene	Working mechanism
PIM-DM	ASM Model	Suitable for small scale, multicast group members relatively dense LAN.	A unidirectional acyclic SPT connecting the multicast source to the group members is maintained through periodic "diffusion-pruning".

Protocol	Model classification	Applicable scene	Working mechanism
PIM-SM	ASM Model	It is suitable for large network with relatively sparse group members and wide distribution.	PIM-SM needs to maintain an RP, set up an RPT, and register a multicast source.
	SSM Model	This scenario is suitable for users in a network who know the location of the multicast source in advance and request multicast data directly from the specified multicast source.	SPT is directly established between the multicast source and group members without the need to maintain RP, build RPT, and register the multicast source.

14.3 Configure PIM-SM

14.3.1 Enable PIM-SM

【Command】

```
ip pim sparse-mode
no ip pim sparse-mode
```

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

`ip pim sparse-mode`: this command is used to enable PIM-SM.

`no ip pim sparse-mode`: this command is used to disable PIM-SM.

By default, the PIM-SM is disabled.

Notice

- This command takes effect only if IP multicast routing is globally enabled.
- For configuration, see the command ip multicast - routing.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim sparse-mode
```

14.3.2 PIM-SM DR Priority

【Command】

```
ip pim sparse-mode dr-priority PRIORITY
no ip pim sparse-mode dr-priority
```

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

PRIORITY: specifies the priority of campaign DR, with the value range from 0 to 4294967294. The higher the value, the higher the priority.

【Description】

`ip pim sparse - mode dr-priority`: this command is used to configure the priority of the campaign DR on the interface.

`no ip pim sparse-mode dr-priority`: this command is used to restore to the default value.

By default, campaign DR has priority of 1.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim sparse-mode dr-priority 10
```

14.3.3 PIM-SM GenID Information

【Command】

ip pim sparse-mode exclude-genid

no ip pim sparse-mode exclude-genid

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip pim sparse-mode exclude-genid: this command is used to configure the interface to send hello messages without GenID information.

no ip pim sparse-mode exclude-genid: this command is used to restore to the default.

By default, the hello message from the interface carries GenID information.

GenID is a random value at the initial creation of an interface to identify unique interface information. With this information, users can detect whether the neighbor device has been restarted.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim sparse-mode exclude-genid
```

14.3.4 PIM-SM Neighbor Reachable State Time

【Command】

ip pim sparse-mode hello-holdtime TIME

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

TIME: specifies the time to maintain the reachable state of the PIM neighbor. The value ranges from 1 to 65535 in seconds. If specified as 65535 seconds, the PIM neighbor is always reachable.

【Description】

ip pim sparse-mode hello-holdtime:this command is used to configure the time on the interface to keep the reachable state of the PIM neighbor.

no ip pim sparse-mode hello-holdtime: this command is used to restore to the default value.

By default, the PIM neighbor is kept in the reachable state for 105 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim sparse-mode hello-holdtime 150
```

14.3.5 Time Interval of PIM-SM Hello Message

【Command】

```
ip pim sparse-mode hello-interval INTERVAL
no ip pim sparse-mode hello-interval
```

【View】

VLAN-IF Interface Configuration View

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

INTERVAL: Specifies the time interval between sending Hello message and the value ranges from 1 to 65535 in seconds.

【Description】

ip pim sparse-mode hello-interval: this command is used to configure the time interval for sending Hello messages on the interface.

no ip pim sparse-mode hello-interval: this command is used to restore to the default value.

By default, Hello messages are sent at an interval of 30 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip pim sparse-mode hello-interval 60
```

14.3.6 PIM-SM Neighbor Filter

【Command】

ip pim sparse-mode neighbor-filter <ACL-NUMBER | ACL-NAME>
no ip pim sparse-mode neighbor-filter

【View】

VLAN-IF Ethernet port configuration mode

Loopback interface configuration view

【Default Level】

2: Configuration level

【Parameter】

ACL-NUMBER: A standard ACL number with values ranging from 1 to 99.

ACL-NAME: SWOS names the ACL name.

【Description】

ip pim sparse-mode neighbor-filter: this command is used to configure an illegal neighbor source address range. The action for the ACL must be deny, and if the action is permit, a mismatch is considered a legitimate neighbor.

no ip pim sparse-mode neighbor-filter: this command is used to cancel configuration of an illegal neighbor source address range.

By default, there are no restrictions on the neighbor source addresses that interfaces can learn from.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 20 deny 192.168.1.0 0.0.0.255
Switch(config)#access-list 20 permit any
Switch(config)#interface vlanif1
```

```
Switch(config-vlanif1)#ip pim sparse-mode neighbor-filter 20
```

14.3.7 Restrictions on Illegal Registration of Messages in PIM-SM

【Command】

```
ip pim accept-register list <ACL-NUMBER | ACL-NAME>
no ip pim accept-register
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

ACL-NUMBER: extend the ACL number, ranging from 100 to 199 and 2000 to 2699.

ACL-NAME: SWOS names the ACL name.

【Description】

ip pim accept-register: this command is used to configure the source address range of the illegal registered message. The action of the ACL must be deny, and the register-stop message will be sent immediately after the match. If it is any other action type, it will be considered as a valid source address.

no ip pim accept-register: This command is used to cancel the configuration of the source address range of illegal registered packets.

By default, there is no restriction on the source address of the interface learnable registered message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 100 deny ip 192.168.1.0 0.0.0.255 any
Switch(config)#access-list 100 permit ip any any
Switch(config)#ip pim accept-register list 100
```

14.3.8 PIM C-BSR

【Command】

ip pim bsr-candidate <IFNAME> [hash-mask-lengh <HASH-LEN> | priority <PRI>]

no ip pim bsr-candidate <IFNAME> [hash-mask-lengh <HASH-LEN> | priority <PRI>]

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

IFNAME: specify the vlanif or Loopback interface for C-BSR.

HASH-LEN: specifies the length of the hash mask, which ranges from 0 to 32.

PRI: specifies the priority of C-BSR, with values ranging from 0 to 255. The higher the value, the higher the priority.

【Description】

ip pim bsr-candidate: this command is used to configure an interface as C-BSR.

no ip pim bsr-candidate: this command is used to remove the C-BSR configuration.

By default, C-BSR is not configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim bsr-candidate vlanif1 hash-mask-length 20
priority 20
```

14.3.9 PIM Registration Message Validation and Compatibility With Cisco Standards

【Command】

ip pim cisco-register-checksum [group-list <ACL-NUMBER | ACL-NAME>]

no ip pim cisco-register-checksum [group-list <ACL-NUMBER | ACL-NAME>]

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

ACL-NUMBER: Standard ACL number with values ranging from 1 to 99 and 1300 to 1999.

ACL-NAME: SWOS names the ACL name.

【Description】

ip pim cisco-register-checksum: this command is used to configure and calculate the verification and compatibility of registered packet with CISCO standard, that is, including the whole PIM mesage, the ACL can be applied to specify the group address range, the action of ACL must be pemit, if no ACL is specified, it means that all multicast groups need to be compatible with CISCO standard.

no ip pim cisco-register-checksum: this command is used to cancel the configure the calculation of the check and compatibility of registered packets with the CISCO standard.

By default, the checksum calculated for the registered message consists of only 8 bytes of the PIM header.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim cisco-register-checksum
```

14.3.10 PIM C-RP is compatible with Cisco standards

【Command】

```
ip pim crp-cisco-prefix
no ip pim crp-cisco-prefix
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip pim crp-cisco-prefix: this command is used to communicate with the BSR of the CISCO device as C-RP.

no ip pim crp-cisco-prefix: this command is used to restore the default configuration.

This feature is not enabled by default.

If the BSR of the CISCO device is not compatible with Prefix Count equal to 0, then if CRP uses the default Group Range, then the Prefix Count in the CRP-ADV message is set to 1, and the Group Range uses 224.0.0.0/4.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim crp-cisco-prefix
```

14.3.11 PIM BSR Ignores C-RP Priority

【Command】

ip pim ignore-rp-set-priority
no ip pim ignore-rp-set-priority

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip pim ignore-rp-set-priority: this command is used to determine whether to ignore the priority of the RP when selecting the RP corresponding to the group.

no ip pim--cisco-prefix: the user restores the default configuration.

This feature is not enabled by default.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim ignore-rp-set-priority
```

14.3.12 The Time Interval Between PIM Sending and/or Pruning Messages

【Command】

ip pim jp-timer VALUE

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

VALUE: Specifies the time interval between sending and adding/pruning messages. The value range is 1 ~ 65535 in seconds.

【Description】

ip pim jp-timer: this command is used to globally configure the time interval for sending join/prune messages.

no Ip pim jp-timer: this command is used to restore to the default value.

By default, the join/prune message is sent at an interval of 60 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim jp-timer 120
```

14.3.13 The Rate at Which PIM Receives and Processes Multicast Service Messages

【Command】

ip pim register-rate-limit VALUE

no ip pim register-rate-limit

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

VALUE: The rate of receiving and processing multicast business message, the value range is 1 ~ 65535, and the unit is one/second.

【Description】

ip pim register-rate-limit: this command is used to globally configure the rate of receiving and processing multicast service messages.

no ip pim register-rate-limit: this command is used to restore to the default value.

By default, there is no limit to the rate of receiving and processing multicast business messages.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim register-rate-limit 1
```

14.3.14 PIM Checks the Accessibility of the RP

【Command】

ip pim register-rp-reachability
no ip pim register-rp-reachability

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip pim register-rp-reachability: this command is used to globally configure whether the RP needs to be checked for reachabability when sending a registration message. If unreachable, CouldRegister is False and CouldRegister cannot be registered.

no ip pim register-rp-reachability: this command is used to restore to the default value.

By default, DR does not need to check the accessibility of PR when sending a registered message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim register-rp-reachability
```

14.3.15 The VLAN interface or source IP address where the PIM sends the registered message

【Command】

```
ip pim register-source < IFNAME | IP-ADDRESS >
no ip pim register-source
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

IFNAME: Vlanif interface for sending registered messages.

IP-ADDRESS: The source IP address used to send the registered message must be the IP address of the interface.

【Description】

ip pim register-source: this command is used to globally configure the VLAN interface or source IP address for sending registered messages.

no ip pim register-source: this command is used to restore to the default values.

By default, the source address of the registered message is the master IP address of the egress interface to the RP in the routing table.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim register-source 192.168.1.100
```

14.3.16 PIM Registration Suppression Time

【Command】

```
ip pim register-suppression VALUE
```

no ip pim register-suppression

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

VALUE: The time interval between receiving the registration stop message and sending the registration message again. The value range is 1 ~ 65535, and the unit is seconds.

【Description】

ip pim register-suppression: this command is used to configure the registered suppression time.

no ip pim register-suppression: this command is used to restore to the default. By default, the registration suppression time is 60 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim register-suppression 120
```

14.3.17 PIM Static PR

【Command】

ip pim rp-address <RP-ADDRESS> [ACL-NUMBER | ACL-NAME]

no ip pim rp-address <RP-ADDRESS> [ACL-NUMBER | ACL-NAME]

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

RP-ADDRESS: Specifies the IP address of the static RP. This address must be a valid unicast IP address and cannot be configured as a 127.0.0.0/8 network segment address.

ACL-NUMBER: Standard ACL number with values ranging from 1 to 99 and 1300 to 1999.

ACL-NAME: SWOS names the ACL name.

【Description】

ip pim rp-address: this command is used to configure the static RP.

no ip pim rp-address: the command is used to delete static RP.

By default, static RP is not configured.

Notice

- When the ACL rule for a static RP reference changes, the RP needs to be re-elected for all multicast groups.
- Repeat this command to configure multiple static RPS. However, if the static RP address or ACL rule specified at the time of configuration is the same, the new configuration overrides the old configuration; If there are multiple static RP services in the same multicast group, then select the static RP with the largest IP address as the group service.
- Dynamic RP has a higher priority than static RP.
- If there are no ACL parameters, the configured static RP will serve all multicast groups (224.0.0.0/4).

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim rp-address 192.168.1.100
```

14.3.18 PIM C-RP

【Command】

ip pim rp-candidate <IFNAME> [priority <PRI> | interval <INTERVAL> | group-list <ACL>]

no ip pim rp-candidate <IFNAME> [priority <PRI> | interval <INTERVAL> | group-list < ACL-NUMBER | ACL-NAME >]

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

IFNAME: Specifies the Vlanif interface of C-RP.

INTERVAL: the sending time interval of CRP-Adv message, the value range is 1 ~ 16383, and the unit is seconds.

PRI: Specifies the priority of C-RP, with values ranging from 0 to 255. The smaller the value, the higher the priority.

ACL-NUMBER: A standard ACL number with values ranging from 1 to 99.

ACL-NAME: SWOS command ACL name.

【Description】

ip pim rp-candidate: this command is used to configure an interface as C-RP.

no ip pim rp-candidate: this command is used to remove the C-RP configuration.

By default, C-RP is not configured.

Notice

- The interface that acts as C-RP must be PIM-SM enabled.
- If C-RP is not specified for the range of multicast groups it serves, the C-RP will serve all multicast groups.
- If the router wants to be a C-RP of multiple group scopes, multiple group scopes need to be represented with multiple ACL rules when configuring the ACL corresponding to group-Listy.
- If this command is executed multiple times for the same interface, the latest configuration overrides the old one.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 20 deny 225.1.1.0 0.0.0.255
Switch(config)#ip pim rp-candidate vlanif1 priority 20 group-list
1 interval 30
```

14.3.19 PIM KAT Timer Aging Time

【Command】

```
ip pim rp-register-kat VALUE
no ip pim rp-register-kat
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

VALUE: The aging time of KAT timer after receiving the registered message, the value range is 1 ~ 65535, and the unit is seconds.

【Description】

ip pim rp-register-kat: the command is used to globally configure the aging time of KAT timer after receiving the registered message.

no ip pim rp-register-kat: the command is used to restore the mode configuration.

By default, after receiving the registration message, the aging time of KAT timer is set to the registration suppression time * 3 + registration detection time (the default is 5 seconds).

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim rp-register-kat 120
```

14.3.20 PIM SPT Switch

【Command】

ip pim spt-threshold [group-list <ACL-NUMBER | ACL-NAME>]
no ip pimspt-threshold [group-list <ACL-NUMBER | ACL-NAME>]

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

ACL-NUMBER: Standard ACL number with values ranging from 1 to 99 and 1300 to 1999.

ACL-NAME: SWOS names the ACL name.

【Description】

ip pim spt-threshold: This command is used to globally configure whether SPT switch can be carried out. ACL can be applied to specify group address range, the action of ACL must be pEMIT. If ACL is not specified, then all multicast groups need to be SPT switch.

no ip pim spt-threshold: this command is used to cancel the recovery configuration.

By default, SPT switching is not performed.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip pim spt-threshold
```

14.3.21 Displays BSR Running Information for PIM-SM

【Command】

```
show ip pim sparse-mode bsr-router
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

View BSR running information for PIM-SM on the current device.

【Instance】

```
Switch#show ip pim sparse-mode bsr-router
PIMv2 Bootstrap information
  BSR address: 172.24.0.10
  Uptime:     00:01:39, BSR Priority: 0, Hash mask length: 30
  Expires:     00:01:31
  Role: Candidate BSR
  State: Candidate BSR

  Candidate RP: 172.24.0.3(vlanif50)
    Advertisement interval 60 seconds
    Next Cand_RP_advertisement in 00:00:28
```

14.3.22 Display PIM-SM Interface Information

【Command】

```
show ip pim sparse-mode interfce [ detail ]
```

【View】

Any view

【Default Level】

1: view level

【Parameter】

interface: specifies the vlanif interface.

detail: specifies to view the details.

【Description】

The command is used to view PIM information on the interface.

【Instance】

```
Switch#show ip pim sparse-mode interface
Address        Interface VIFindex  Ver/    Nbr    DR    DR
                                   Mode   Count  Prior
10.3.101.252        vlanif10    0            v2/S     1          1
10.3.101.253
11.1.1.2         vlanif11  2       v2/S   1      1     11.1.1.3
172.24.0.2       vlanif50  3       v2/S   1      1     172.24.0.3

Switch#show ip pim sparse-mode interface detail
vlanif10 (vif 0):
  Address 10.3.101.252, DR 10.3.101.253
  Hello period 30 seconds, Next Hello in 2 seconds
  Triggered Hello period 5 seconds
  Neighbors:
   10.3.101.253

vlanif11 (vif 2):
  Address 11.1.1.2, DR 11.1.1.3
  Hello period 30 seconds, Next Hello in 2 seconds
  Triggered Hello period 5 seconds
  Secondary addresses:
   11.1.1.1
  Neighbors:
   11.1.1.3

vlanif50 (vif 3):
  Address 172.24.0.2, DR 172.24.0.3
  Hello period 30 seconds, Next Hello in 3 seconds
  Triggered Hello period 5 seconds
  Neighbors:
   172.24.0.3
```

14.3.23 Displays the Local Multicast Group Members of the PIM-SM

【Command】

show ip pim sparse-mode local-members [<IFNAME>]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: Specifies the vlanif interface.

【Description】

View the local multicast group members of PIM-SM of the current device.

【Instance】

```
Switch#show ip pim sparse-mode local-members vlanif10
PIM Local membership information
vlanif10:
  (*, 225.2.2.2) : Include
```

14.3.24 Displays the PIM-SM Multicast Routing Table Entry

【Command】

show ip pim sparse-mode mroute [GROUP-ADDRESS] [detail]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

GROUP-ADDRESS: Specifies the multicast group address.

detail: specifies to view the details.

【Description】

View the current PIM-SM multicast routing table entry.

【Instance】

```
Switch#show ip pim sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 1
(S,G,rpt) Entries: 1
FCR Entries: 0

(10.3.101.111, 239.255.255.250)
RPF nbr: 0.0.0.0
RPF idx: None
SPT bit: 1
Upstream State: NOT JOINED
 Local    ..............................
 Joined   ..............................
 Asserted ..............................
 Outgoing ..............................

(10.3.101.111, 239.255.255.250, rpt)
RP: 0.0.0.0
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: RPT NOT JOINED
 Local    ..............................
 Pruned   ..............................
 Outgoing ..............................

*DUT4#show ip pim sparse-mode mroute detail
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 1
(S,G,rpt) Entries: 1
FCR Entries: 0

(10.3.101.111, 239.255.255.250) Uptime: 00:09:05 Flags: 0x03
  RPF nbr: None, RPF idx: None
  Upstream:
   State: NOT JOINED, SPT Bit: on, JT: off, KAT Expiry: 85 secs
```

```
  Downstream:

(10.3.101.111, 239.255.255.250, rpt) Uptime: 00:09:05
  RP: 0.0.0.0, RPF nbr: None, RPF idx: None
  Upstream:
   State: RPT NOT JOINED, OT: off
   Macro state:
  Downstream:
```

14.3.25 Display PIM-SM Neighbor Information

【Command】

show ip pim sparse-mode neighbor [<IFNAME> | detail]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: Specifies the vlanif interface

detail: specifies to view the details.

【Description】

View the currently learning PIM-SM neighborhood running status.

【Instance】

```
Switch#show ip pim sparse-mode neighbor
Neighbor        Interface         Uptime/Expires    Ver   DR
Address                                             Priority/Mode
10.3.101.253      vlanif10          01:53:56/00:01:17  v2    1 /
DR
11.1.1.3         vlanif11          01:53:57/00:01:18  v2    1 / DR
172.24.0.3       vlanif50          01:53:56/00:01:17  v2    1 / DR
```

14.3.26 Display PIM-SM Next Hop Information

【Command】

show ip pim sparse-mode nexthop

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

View the next hop information used by the PIM-SM module in the current device.

【Instance】

```
Switch#show ip pim sparse-mode nexthop
Flags: N = New, R = RP, S = Source, U = Unreachable
Destination  Type  Nexthop   Nexthop      Nexthop  Nexthop Metric
Pref  Refcnt
               Num     Addr         Ifindex    Name
_____________________________________________________________________
_____________
10.3.101.111  ..S.  1         0.0.0.0         4                0
0     2
172.24.0.10   .R..  1         0.0.0.0         6                0
0     1
```

14.3.27 Display PIM-SM RP Information

【Command】

show ip pim sparse-mode rp mapping

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

View the RP-Set information for the PIM-SM in the current device.

【Instance】

```
Switch#show ip pim sparse-mode rp mapping
PIM Group-to-RP Mappings
Group(s): 224.0.0.0/4
  RP: 172.24.0.10
```

```
    Info source: 172.24.0.10, via bootstrap, priority 192
        Uptime: 00:08:35, expires: 00:01:55
  RP: 172.24.0.3
    Info source: 172.24.0.10, via bootstrap, priority 192
        Uptime: 00:08:35, expires: 00:01:55
  RP: 172.24.0.2
    Info source: 172.24.0.10, via bootstrap, priority 192
        Uptime: 00:08:35, expires: 00:01:55
Group(s): 224.0.0.0/4, Static
   RP: 172.24.0.10
        Uptime: 02:28:07
```

14.3.28 Displays the RP Address of the PIM-SM Multicast Group

【Command】

show ip pim sparse-mode rp-hash <GROUP-ADDRESS>

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

GROUP-ADDRESS: Specifies the multicast group address.

【Description】

View the RP address for the specified multicast group address.

【Instance】

```
Switch#show ip pim sparse-mode rp-hash 226.1.1.1
   RP: 172.24.0.2
   Info source: 172.24.0.10, via bootstrap
```

15 VRRP Configuration

15.1 Overview

The Virtual Router Redundancy Protocol (VRRP) groups multiple routing devices into a virtual router and uses the virtual gateway device's IP address as the default gateway address. When the gateway fails, VRRP selects a new gateway to transmit service traffic to ensure reliable communication.

As networks rapidly develop and applications become diversified, various value-added services such as Internet Protocol television (IPTV) and video conferencing services are widely deployed. Demands for network infrastructure reliability are increasing, especially for nonstop service transmission.

Generally, all hosts on the same network segment are configured with the same default route with the gateway address as the next hop address. Messages sent by the host to other network segments will be sent to the gateway through the default route, and then forwarded by the gateway, thus realizing the communication between the host and the external network. When the gateway fails, all hosts in this network segment that use the gateway as the default route will not be able to communicate with the external network. A common method to improve network reliability is to configure multiple egress gateways. However, route selection between the gateways becomes an issue.

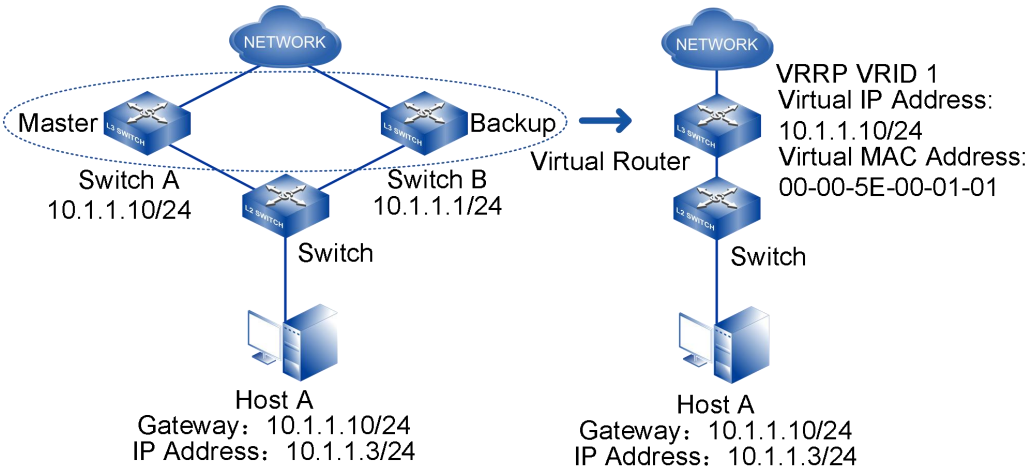
VRRP resolves this issue. VRRP virtualizes multiple routing devices into a virtual router without changing the networking. The virtual router IP address is configured as the default gateway address. When the gateway fails, VRRP selects a new gateway to transmit service traffic to ensure reliable communication.

On a multicast or broadcast LAN, such as Ethernet, VRRP provides a highly reliable link when the gateway fails, without modifying host and gateway configurations. VRRP prevents network interruptions when a single link fails.

15.2 Principle Description

15.2.1 Concepts

As shown in the following figure, HostA is dual-homed to SwitchA and SwitchB through the switch. SwitchA and SwitchB constitute a VRRP group to implement link redundancy.



The VRRP protocol can be deployed in the network as shown in the figure above. The basic concepts of the VRRP protocol are introduced in combination with the figure below:

- VRRP router: device running VRRP. It may join one or more virtual routers. SwitchA and SwitchB are VRRP routers.
- Virtual router: VRRP group consisting of one master and multiple backups. The VRRP group's virtual IP address is used as the default gateway address on a LAN. SwitchA and SwitchB constitute a virtual router.
- Virtual router master: VRRP device that forwards packets. SwitchA is the virtual router master.
- Virtual router backup: a group of VRRP devices that do not forward packets. When the master is faulty, a backup with the highest priority becomes the master. SwitchB is the virtual router backup.
- VRID: virtual router ID. The VRID of the virtual router composed of SwitchA and SwitchB is 1.
- Virtual IP Address: the IP address of a virtual router. A virtual router can have one or more IP addresses, which can be configured by the user. For example, the virtual IP address of the virtual router composed of SwitchA and SwitchB is 10.1.1.10/24.
- IP address owner: VRRP device that uses an IP address of a virtual router as the actual interface address. If an IP address owner is available, it usually functions

as the virtual router master. The interface address of SwitchA and the IP address of the virtual router are both 10.1.1.10/24, so SwitchA is the IP address owner.

- Virtual MAC address: MAC address that is generated by the virtual router based on the VRID. A virtual router has one virtual MAC address and is in the format of 00-00-5E-00-01-{VRID} (VRRP for IPv4) or 00-00-5E-00-02-{VRID} (VRRP for IPv6). The virtual router sends ARP Reply packets carrying the virtual MAC address but not the interface MAC address. The VRID of the virtual router composed of SwitchA and SwitchB is 1, so the MAC address of the VRRP group is 00-00-5E-00-01-01.

15.2.2 VRRPv2 and VRRPv3 Advertisement Packets

The master sends VRRP Advertisement packets to notify all backups in the VRRP group of the master's priority and status.

VRRP Advertisement packets are encapsulated into IP packets and sent to the VRRP virtual IP address. In the IP packet header, the source address is the primary IP address of the interface that sends the packets (not the virtual IP address), the destination address is 224.0.0.18, the TTL is 255, and the protocol number is 112.

The primary IP address is selected from one of actual IP addresses of interfaces. Usually, it is the first configured IP address.

VRRP has two versions: VRRPv2 and VRRPv3. VRRPv2 applies to only the IPv4 network, and VRRPv3 applies to IPv4 and IPv6 networks.

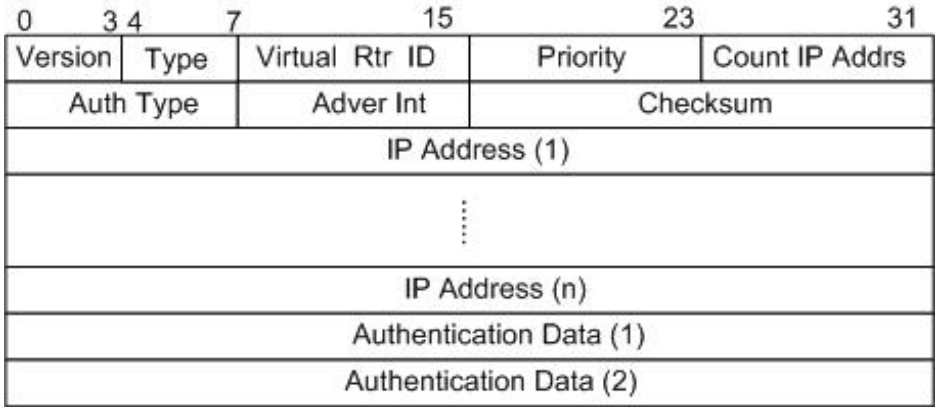
VRRP is classified into VRRP for IPv4 (VRRP) and VRRP for IPv6 (VRRP6) by network type. VRRP supports VRRPv2 and VRRPv3, and VRRP6 supports only VRRPv3.

At present, our equipment only supports VRRPv2 version.

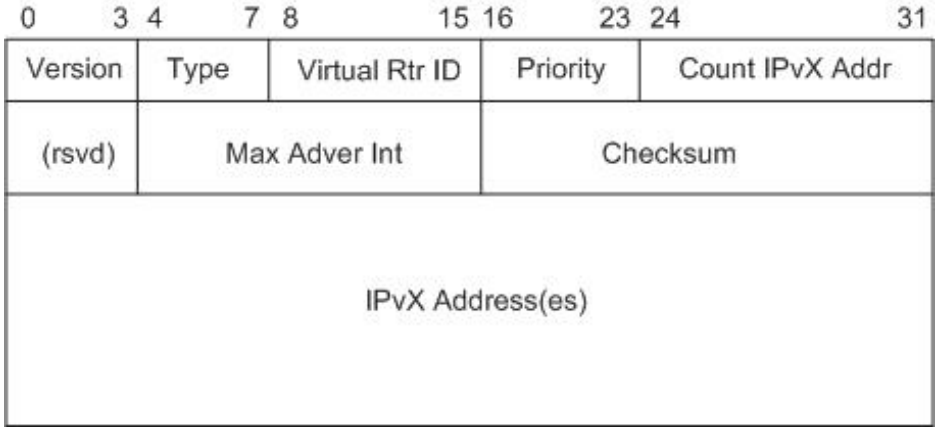
15.2.2.1 VRRPv2 and VRRPv3 Advertisement Packet Formats

The message structures of VRRPv2 and VRRPv3 are shown in the following figure.

- Message structure of VRRPv2:



- Message structure of VRRPv3:



The meanings of each field are shown in the following table:

Field	Definition	
	VRRPv2	VRRPv3
Version	VRRP protocol version. The value is 2.	VRRP protocol version. The value is 3.
Type	VRRP Advertisement packet type. The value 1 indicates an Advertisement packet.	VRRP Advertisement packet type. The value 1 indicates an Advertisement packet.
Virtual Rtr ID (VRID)	Virtual router ID. The value ranges from 1 to 255.	Virtual router ID. The value ranges from 1 to 255.
Priority	Priority of the master in a VRRP group. The value ranges from 0 to 255. The value 0 indicates that the device stops participating in the VRRP group	Priority of the master in a VRRP group. The value ranges from 0 to 255. The value 0 indicates that the device stops participating in the VRRP group

Field	Definition	
	VRRPv2	VRRPv3
	so that the backup with the highest priority can become the master immediately. The value 255 is reserved for the IP address owner. The default value is 100.	so that the backup with the highest priority can become the master immediately. The value 255 is reserved for the IP address owner. The default value is 100.
Count IP Addrs/Count IPvX Addr	Number of virtual IPv4 addresses in the VRRP group.	Number of virtual IPv4 or IPv6 addresses in the VRRP group.
Auth Type	Authentication mode. There are three authentication modes: • 0: Non Authentication • 1: Simple Text Password • 2: IP Authentication Header (MD5 authentication)	-
Adver Int/Max Adver Int	Interval at which VRRP Advertisement packets are sent, in seconds. The default value is 1.	Interval at which VRRP Advertisement packets are sent, in centiseconds. The default value is 100 (1 second).
Checksum	16-bit checksum, which is used to verify data integrity in VRRP Advertisement packets.	16-bit checksum, which is used to verify data integrity in VRRP Advertisement packets.
IP Address/IPvX Address(es)	Virtual IPv4 address in the VRRP group. The Count IP Addrs field determines the number of virtual IPv4 addresses in the VRRP group.	Virtual IPv4 or IPv6 address in the VRRP group. The Count IPvX Addrs field determines the number of virtual IPv4 or IPv6 addresses in the VRRP group.
Authentication Data	Authentication key. This field is used only in simple authentication and MD5 authentication modes. In other authentication modes, this field	-

Field	Definition	
	VRRPv2	VRRPv3
	is filled with 0s.	
rsvd	-	Reserved. This field has a fixed value of 0.

VRRPv2 and VRRPv3 have the following differences:

- Apply to different networks. VRRPv3 applies to IPv4 and IPv6 networks, whereas VRRPv2 applies to only IPv4 networks.
- Have different authentication functions. VRRPv3 does not support authentication, whereas VRRPv2 supports authentication.
- Use different time units to measure the interval at which VRRP Advertisement packets are sent. VRRPv3 uses centiseconds, whereas VRRPv2 uses seconds.

15.2.2.2 VRRP Authentication

Different authentication modes and authentication keys can be set in VRRPv2 Advertisement packets:

- Non-authentication: The local device does not authenticate VRRP Advertisement packets before sending them. The remote device does not authenticate the received VRRP Advertisement packets and considers all the received packets valid.
- Simple authentication mode: the device sending VRRP notification message fills the authentication mode and authentication word into the notification message, while the device receiving the notification message matches the authentication mode and authentication word in the message with the authentication mode and authentication word configured by the local terminal. If they are the same, it is considered that the received message is a legal VRRP notification message. Otherwise, the received message is considered as an illegal message, and the message is discarded.
- MD5 authentication: The local device uses the MD5 algorithm to encrypt the authentication key and encapsulates the key in the Authentication Data field of an outgoing VRRP Advertisement packet. Upon receipt of the VRRP Advertisement packet, the remote device decrypts the authentication key and checks whether the authentication mode and authentication key are the same as those configured locally. The remote device then accepts or discards the packet depending on the authentication result.

15.2.3 VRRP Implementation

15.2.3.1 VRRP State Machine

VRRP defines three statuses: Initialize, Master, and Backup. Only the device in Master state can forward packets destined for the virtual IP address.

State	Note
Initialize	VRRP is unavailable. The device in Initialize state cannot process VRRP Advertisement packets. When VRRP is configured on the device or the device detects a fault, it enters the Initialize state. After receiving an interface Up message, the device whose priority is 255 becomes the master. The device whose priority is less than 255 switches to the Backup state.
Master	The VRRP device in Master state performs the following operations: • Sends VRRP Advertisement packets at intervals. • Uses the virtual MAC address to respond to ARP Request packets destined for the virtual IP address. • Forwards IP packets destined for the virtual MAC address. • And if it is the owner of the virtual IP address, receiving an IP message with the destination IP address as the virtual IP address. Otherwise, discard the IP message. • Becomes the backup if the device receives a VRRP Advertisement packet with a higher priority than its VRRP priority. • Becomes the backup if the device receives a VRRP Advertisement packet with the same priority as its VRRP priority and the IP address of the local interface is smaller than that of the connected interface on the remote device.
Backup	The VRRP device in Backup state performs the following operations: • Receives VRRP Advertisement packets from the master and determines whether the master is working properly. • Does not respond to ARP Request packets destined for the virtual IP address. • Processes the IP packets destined for the virtual IP address

State	Note
	based on the Layer 2 forwarding process. • When receiving a packet of lower priority, it immediately switches to the Master state by default. If non-preemption is configured, the device resets the timer. If a preemption delay is configured, the device resets the timer and switches to the Master state after the preemption delay expires. When receiving a packet of higher priority, the device resets the timer. When receiving a packet of equal priority, the device resets the timer but does not compare IP addresses. Master_Down_Interval timer: If the backup does not receive Advertisement packets after the timer expires, the backup becomes the master. The formula is as follows: master _ down _ interval = (3 * advertisement _ interval)+skew _ time. With skew _ time = (256–priority)/256. • Master_Down_Interval = 3xAdvertisement_Interval + Skew_time (offset time)

15.2.3.2 Working process of VRRP

The working process of VRRP is as follows:

1 The equipment in VRRP backup group elects Master according to priority. Master device sends free ARP message to inform the device or host connected with it of the virtual MAC address, so as to undertake the message forwarding task.

2 The VRRP working process is as follows:

3 Devices in a VRRP group select the master based on their priorities. The master sends gratuitous ARP packets to notify the connected network devices or hosts of the virtual MAC address of the VRRP group.

4 The master periodically sends VRRP Advertisement packets to all backups in the VRRP group to advertise its configuration (for example, priority) and running status.

5 When the original Master device fails to recover, if the device is the IP address owner (priority is 255), it will be directly switched to the Master state. If the device has a priority less than 255, it is first switched to the Backup state, and its priority is restored to the priority configured before the failure.

6 If the original master is replaced by another device in the group, the new master sends gratuitous ARP packets carrying the virtual MAC address and virtual IP

address of the virtual router. The new master uses these packets to update the MAC address entry on the connected network devices or hosts. User traffic is then switched to the new master. This process is transparent to users.

- When the original master recovers and is the IP address owner (with priority 255), the original master switches to the Master state. If the priority of the original master is smaller than 255, the device switches to the Backup state. The priority of the device is then restored to the original value before the failure.
- If the backup has a higher priority than the master, the working mode of the backup (preemption or non-preemption) determines whether the master is re-selected.

Preemption mode: If the priority of a backup is higher than the priority of the current master, the backup automatically becomes the master.

- Non-preemption mode: As long as the master is working properly, the backup with a higher priority cannot become the master.
- To ensure that the master and backup work properly, VRRP must be able to select the master and advertise the master status.

The detailed VRRP working process is as follows:

- **Selecting the master**

  VRRP determines the device role in the virtual router based on device priorities. The device with a higher priority is more likely to become the master.

  The VRRP device initially created works in Initialize state. after receiving the message of interface Up, if the priority of the device is 255, it will directly become the Master device; If the priority of the device is less than 255, it will switch to the Backup state first, and then switch to the Master state after the Master_Down_Interval timer expires. At first, the VRRP device switched to the Master state obtains the priority of other members in the virtual device through the interaction of VRRP notification messages, and carries out the Master election:

  - If the master priority in VRRP Advertisement packets is higher than or equal to the priority of the device, the backup remains in Backup state.
  - If the master priority in VRRP Advertisement packets is lower than the priority of the device, the backup switches to the Master state in preemption mode or remains in Backup state in non-preemption mode.
  - If multiple devices in the VRRP group switch to the master, the devices with a lower priority switch to the Backup state. The device with the highest priority becomes the master after these devices exchange VRRP

Advertisement packets. If multiple devices have the same priority, the device where the interface with the largest IP address resides is the master.

- If the device is the IP address owner, it switches to the Master state immediately after receiving an interface Up message.

- **Advertising the master status**

  The master periodically sends VRRP Advertisement packets to all backups in the VRRP group to advertise its configuration (for example, priority) and running status. The backup determines whether the master works properly based on the received VRRP Advertisement packets.

  - When the Master device voluntarily gives up the Master status (for example, the Master device quits the Backup group), it will send an announcement message with priority 0 to make the backup device switch to the Master device quickly without waiting for the Master_Down_Interval timer to expire. This switching time is called Skew time, and its calculation method is: (priority of 256－Backup device) /256, and its unit is seconds.
  - If the master cannot send VRRP Advertisement packets due to network faults, the backups cannot learn the running status of the master immediately. The backups consider the master faulty only after the Master_Down_Interval timer expires. Then a backup switches to the Master state. Master_Down_Interval = 3 x Advertisement_Interval + Skew_time (in seconds)

Note

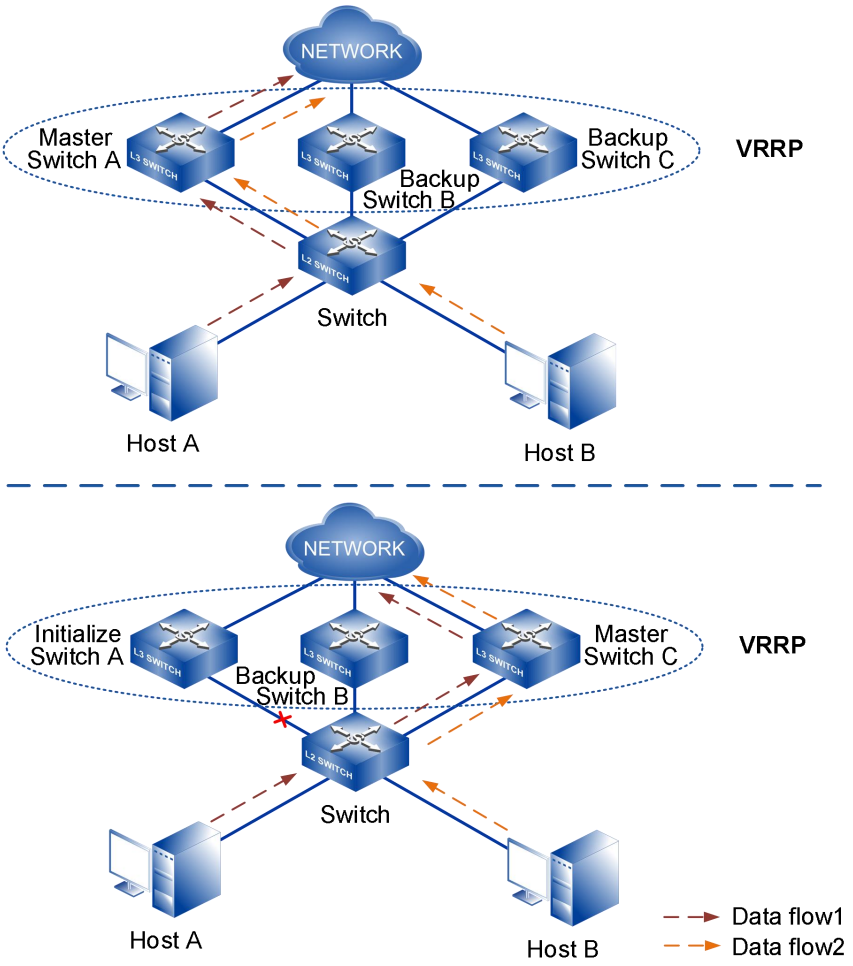
If congestion occurs on an unstable network, the backup may not receive VRRP Advertisement packets from the master within the period of Master_Down_Interval. A backup then switches to the Master state. If the VRRP Advertisement packet from the original master reaches the backup (new master), the new master switches to the Backup state. In this case, the VRRP group status changes frequently. To solve the problem, the preemption delay is used. When the Master_Down_Interval timer expires, the backup waits for the preemption delay. If the backup does not receive a VRRP Advertisement packet within the preemption delay, it switches to the Master state.

15.2.4 VRRP Active/Standby

Master backup is the basic way of VRRP to provide backup function, as shown in the following figure. In active/standby mode, a virtual router consists of one master and multiple backups.

Under normal circumstances, SwitchA is the Master device and undertakes business forwarding tasks, while SwitchB and SwitchC are Backup devices and do not undertake business forwarding. SwitchA regularly sends VRRP notification messages to inform SwitchB and SwitchC that they are working normally. In case of failure of SwitchA, SwitchB and SwitchC will elect new Master devices according to priority, continue to forward data for the host, and realize the function of gateway backup.

After the failure of SwitchA is restored, it will be re-elected as Master； in the preemption mode; In the non-preemptive mode, it will remain in the Backup state.



The following figure briefly illustrates the basic principle of VRRP master backup.Active/Standby is the basic VRRP working mode, as shown in Figure.

- SwitchA is the master, with the priority 120. It uses the delayed preemption mode.

- SwitchB is the backup, with the priority 100. It uses the immediate preemption mode.
- SwitchC is the backup, with the priority 110. It uses the immediate preemption mode.

1 When SwitchA is running properly, traffic sent from users is transmitted along the path Switch -> SwitchA -> Router. SwitchA periodically sends VRRP Advertisement packets to SwitchB and SwitchC, notifying them that the master is working properly.

2 When a fault occurs on SwitchA, VRRP does not function on SwitchA. Because SwitchC has a higher priority than SwitchB, SwitchC becomes the master. SwitchC starts to send VRRP Advertisement packets and gratuitous ARP packets, whereas SwitchB remains in Backup state. User traffic is transmitted along the path Switch -> SwitchC -> Router.

3 When SwitchA recovers, its priority is restored to 120 and it enters the Backup state. SwitchC continues sending VRRP Advertisement packets. When SwitchA receives a VRRP Advertisement packet, it compares the priority in the packet with its own priority and detects that its priority is higher. After the preemption delay, SwitchA becomes the master and starts to send VRRP Advertisement packets and gratuitous ARP packets. User traffic is again transmitted along the path Switch -> SwitchA -> Router.

15.2.5 VRRP Load Balancing

In load balancing mode, multiple devices transmit service traffic simultaneously. Therefore, the load balancing mode requires two or more virtual routers. Each virtual router contains one master and multiple backups, and the master in each virtual router can be different.

VRRP active/standby and load balancing modes are similar in terms of implementation and packet negotiation processes. The load balancing mode differs from the active/standby mode in the following ways:

- Multiple VRRP groups need to be created, and the master in each VRRP group can be different.
- A VRRP device can join multiple VRRP groups and has different priorities in different VRRP groups.

15.2.6 VRRP Monitoring the Uplink Status

Additional technologies are required to enhance the VRRP active/standby function. For example, if the link from the master to a network is disconnected, VRRP cannot detect the fault or switch services. In addition, hosts cannot remotely access the network through the master. To solve this problem, configure VRRP to monitor the specified interface or uplink status.

When the master detects that the interface or uplink fails, the master reduces its priority to be lower than that of the backup and immediately sends VRRP Advertisement packets. After the backup receives the VRRP Advertisement packets, it detects that the priority in the VRRP Advertisement packets is lower than its priority. The backup switches to the master. This ensures traffic forwarding.

15.3 Configure VRRP

15.3.1 Create VRRP Group and Enter VRRP Configuration View

【Command】

router vrrp VIRTUAL-ROUTER-ID

no router rip VIRTUAL-ROUTER-ID

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

Virtual-router-id: VRRP backup group number, value range is 1~255.

【Description】

router vrrp: this command is used to create a VRRP backup group session and enter session configuration mode.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
```

15.3.2 VRRP Interface Enablement

【Command】

network interface <IFNAME>

no network interface <IFNAME>

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

IFNAME: interface name, only vlanif interface is supported.

【Description】

network interface: this command is used to enable VRRP on the specified network segment interface.

By default, VRRP function on the interface is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch (config)#router vrrp 1
Switch (config- router)#network interface vlanif1
```

15.3.3 VRRP Virtual IP Address

【Command】

virtual-ip <A.B.C.D > (master | backup)

no virtual-ip

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: VRRP group virtual address.

Master: set the default master device, that is, the IP address owner.

Backup: set the default backup device.

【Description】

virtual-ip: this command is used for the virtual IP address of the VRRP group, which should be on the same subnet with the primary address of the interface specified by the network command.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#virtual-ip 192.168.1.10 backup
```

15.3.4 Preemption Mode Enablement

【Command】

preempt-mode （true | false）

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

true: enable preemption mode.

false: disable preemption mode.

【Description】

preempt-mode: this command is used to set the switch in the backup group to work in preemption mode. By default, the preempt-delay-time is 0 second and can be set by the preempt-delay-time command.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#preempt-mode true
```

15.3.5 Preemption Delay Time

【Command】

preempt-delay-time DELAY-VALUE

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

DELAY-VALUE: preempt delay time, the value range is 0~255, the unit is seconds, the default value is 0 seconds.

【Description】

preempt-delay-time: this command is used to set the preemption delay for the VRRP backup group. In order to avoid frequent master/backup state transitions of the backup group members and make backup router have enough time to collect the necessary information (such as routing information), when backup router receives notification message with lower priority than the local priority, instead of immediately preempting to become Master, it will wait for a certain time - after the preemption delay time, VRRP notification message will be sent out to replace the original Master router.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#preempt-delay-time 5
```

15.3.6 Priority of VRRP Device

【Command】

```
priority PRIORITY-VALUE
no priority
```

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

PRIORITY-VALUE: the value of priority, ranging from 1 to 255. The larger the value, the higher the priority. If it is the owner of the virtual IP address, then the priority must be 255.

【Description】

priority: this command is used to set the priority of the router in the backup group.

no priority: this command is used to restore to the default value.

By default, the router has a priority of 100 in the backup group.

Notice

- Priority determines the router's status in the backup group, and the higher the priority, the more likely it is to become a Master router. Priority 0 is reserved by the system for special use and 255 is reserved by the system for IP address owners.
- If the router is the owner of the IP address, its priority is always 255, indicating that it is a Master router as long as it is working properly.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#priority 150
```

15.3.7 VRRP Notification Message Interval

【Command】

advertisement-interval ADVER-INTERVAL

no advertisement-interval

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

ADVER-INTERVAL: the interval between the Master router in the backup group sending VRRP notification messages, with a value range of 1~10 in seconds.

【Description】

advertisement-interval: this command is used to set the time interval between VRRP notification messages sent by the Master router in the backup group.

no advertisement-interval: this command is used to restore to the default value.

By default, the Master router in the backup group sends VRRP notification messages every second.

Notice

Routers within the same backup group are configured at the same time interval.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#advertisement-interval 5
```

15.3.8 VRRP Message Authentication Mode

【Command】

```
authentication-mode （md5 | text）
no authentication-mode
```

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

Text: means authentication method is simple character authentication.

md5: md5 algorithm for authentication, not currently supported.

【Description】

authentication-mode: this command is used to configure how the backup group sends and receives VRRP messages.

no authentication-mode: this command is used to restore to the default value.

By default, authentication is not performed.

Notice

Different backup groups on an interface can set different authentication mode and authentication word; Members joining the same backup group need to set the same authentication mode and authentication word.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#authentication-mode text
```

15.3.9 Authentication Word of VRRP Message

【Command】

```
authentication-string KEY
no authentication-string
```

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

KEY: authentication word, case sensitive

【Description】

authentication-string: this command is used to configure how the backup group sends and receives authentication word of VRRP messages.

no authentication-string: this command is used to restore to the default value. By default, authentication is not performed.

Notice

Different backup groups on an interface can set different authentication mode and authentication word; Members joining the same backup group need to set the same authentication mode and authentication word.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#authentication-string vrrpkey
```

15.3.10 Configure the Monitoring Specified Interface

【Command】

circuit-failover IFNAME PRIORITY-DELTA

no circuit-failover

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

IFNAME: interface name, physical ports and vlanif interfaces is supported.

PRIORITY -DELTA: amount of priority reduction, ranging from 1 to 253.

【Description】

circuit-failover: this command is used to configure to monitor the specified interface.

no circuit-failover: this command is used to cancel the monitor of the specified interface.

By default, no monitored interface is specified.

When the interface connecting the router to the uplink fails, the backup group cannot perceive the uplink failure. If the router is in the Master state at this time, the host in the LAN cannot access the external network. This problem can be solved by monitoring the function of the specified interface. When the interface connected to the upline is in the Down state, the router actively lowers its priority, making the priority of other routers in the backup group higher than that of this router, so that the router with the highest priority becomes Master and undertakes the forwarding task.

Notice

When the status of the monitored interface changes from Down to Up, the priority level of the corresponding router will be automatically restored.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#circuit-failover ge2 10
```

15.3.11 Track IPDT Session

【Command】

track ipdt session SESSION-ID (increased | reduced) PRIORITY-DELTA

no track ipdt session SESSION-ID

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

SESSION-ID: ipdt session ID, ranging from 1 to 8.

increased: upgrade priority, usually used for backup devices.

reduced: reduce priority, usually used for master devices.

PRIORITY-DELTA: amount of priority upgrade or reduction, ranging from 1 to 253.

【Description】

track ipdt session: this command is used to set up sessions to track ipdt.

No track ipdt session: this command is used to cancel sessions to track ipdt.

By default, no traced ipdt session is specified.

Similar to circuit-failover, the VRRP backup group tracks an ipdt session, and with the up or down of the ipdt session, the priority is increased or decreased accordingly. Usually reduced is set on the master device and increased is set on the backup device.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#track ipdt session 1 increased 20
```

15.3.12 Enable VRRP Backup Group

【Command】

enable

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

enable: this command is used to enable the VRRP backup group, which is not enabled by default.

Notice

- When modifying any configuration of the backup group, enable must be turned off first.
- The backup group must be configured with the virtual-ip and network interface commands to enable successfully.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
Switch(config- router)#enable
```

15.3.13 Disable VRRP Backup Group

【Command】

disable

【View】

VRRP backup group session view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

disable : this command is used to disable the enabled VRRP backup group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router vrrp 1
```

```
Switch(config- router)#disable
```

15.3.14 Display VRRP Backup Group Information

【Command】

```
show vrrp [ virtual-router-id ]
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

VITUAL-ROUTER-ID: backup group, the value range is 1~8.

【Description】

View the current configuration and operation of the specified VRRP backup group or all VRRP backup groups.

【Instance】

```
Switch> enable
Switch#show vrrp
VrId <10>
 State is Master
 Send Adv after 1 seconds
 Virtual IP is 11.1.1.1 (Not IP owner)
 Interface is vlanif11
 Priority is 151
 Advertisement interval is 1 sec
 Preempt mode is TRUE
 Preempt delay time is 30 sec
```

16 RIP Configuration

16.1 Overview

RIP is a distance-vector protocol that uses hop count to measure the distance to a destination. It is a simple Interior Gateway Protocol (IGP) that is easier to implement, configure, and manage than other routing protocols such as OSPF and IS-IS. RIP exchanges routing information using UDP packets through UDP port 520.

Two RIP versions are used in IPv4 networks: RIP version 1 (RIP-1) and RIP version 2 (RIP-2). RIP-2 is an extension to RIP-1.

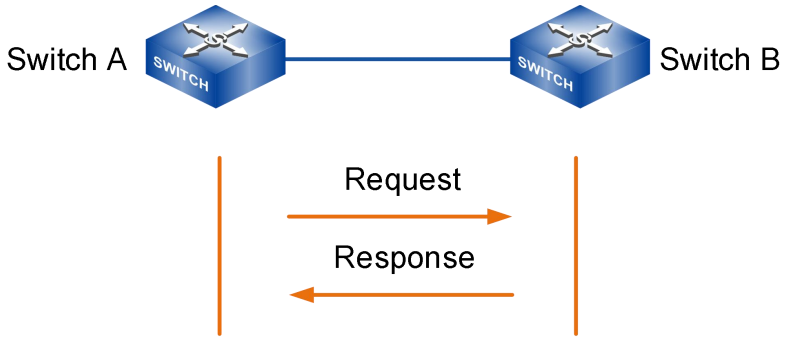
Because RIP is relatively simple to implement, configuration and maintenance management is far easier than OSPF and IS-IS, so RIP is mainly used in small-sized networks, such as campus networks and regional networks with simple structure. RIP isn't used in more complex environment and large network.

16.2 Principles

16.2.1 RIP Principles

16.2.1.1 RIP Routing Table

When RIP starts on a switch, the RIP routing table on the switch contains only direct routes. Neighboring switches on different network segments can communicate with each other only after they learn routing entries from each other.



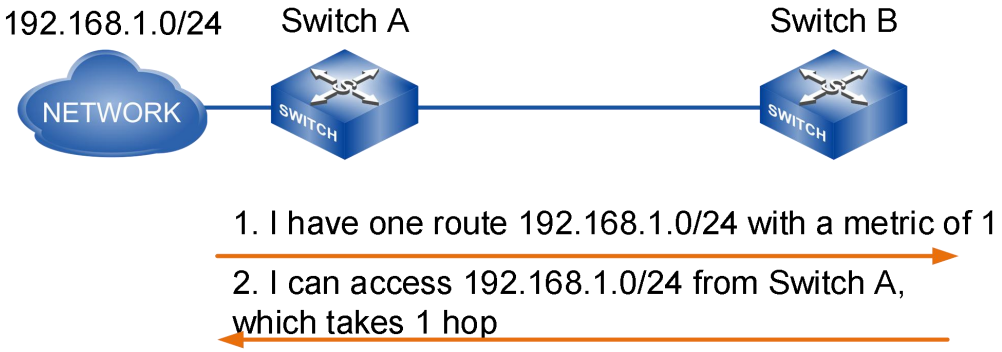
The formation process of RIP routing is shown in the figure above.

4 Once the RIP protocol is activated, the SwitchA broadcasts a Request message to the neighboring switch.

5 SwitchB receives the Request packet, encapsulates its RIP routing table into a Response packet, and then broadcasts the Response packet to the network segment connected to the interface that received the Request packet.

6 SwitchA receives the Response packet and then generates its RIP routing table based on this packet.

RIP updates and selects routes through route advertisements. In this situation, switches do not know the entire network topology and just know the distances to destination networks and the direction or interface used to reach destination networks. As shown in the figure below, SwitchB received a routing announcement from SwitchA. At this point, SwitchB knows that a 192.168.1.0/24 network can be reached through a SwitchA, with a measure of 1 hop, and no other information is known to SwitchB. Even though information contained in this advertisement becomes incorrect, SwitchB still considers that it can reach the network 192.168.1.0/24 through SwitchA with the hop count 1. This is the root cause of RIP networks prone to routing loops.

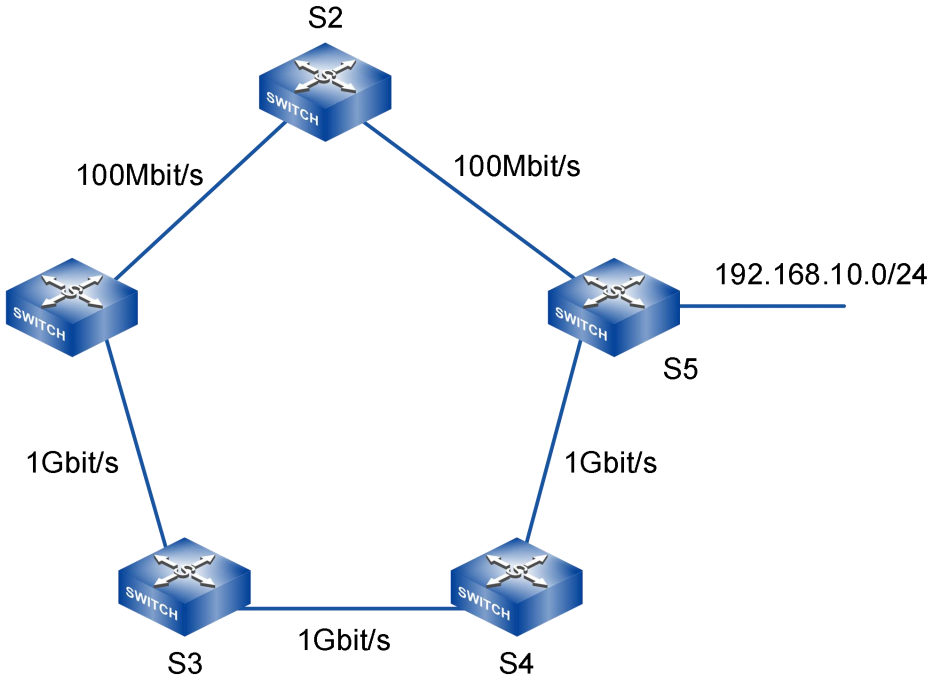


16.2.1.2 RIP Metric

In RIP, the default hop count from a device to its directly connected network is 0, and the hop count from a device to a reachable network through another device is 1. That is, the hop count (the metric) equals the number of devices along the path from the local network to the destination network.

As shown in the figure below, there are two paths from S1 to 192.168.10.0/24 network segment:

- S1-s2-s5 passes through two devices S2 and S5 in the middle. The measurement value of this path is 2 hops and the bandwidth is 100Mbit/s.
- S1-s3-s4-s5, through the middle of S3, S4, S5 three devices, the measurement value of the path is 3 hops, but the bandwidth is large, is a Gigabit link.



As required by the RIP metric standards, packets will be forwarded along the path passing through S2; however, the link bandwidth of this path is not the highest. This causes low bandwidth efficiency on the RIP network and hinders QoS management.

To prevent the hop count from becoming infinite when RIP routes are flooded indefinitely on the network, and to limit the convergence time, RIP allows a maximum of 15 hops. A hop count of 16 or greater is defined as infinity, making the destination network or host unreachable. This design limits the network scale that RIP supports. Therefore, RIP is not suitable for large-scale networks.

16.2.1.3 RIP Route Update and Maintenance

RIP protocol mainly USES three timers to update and maintain routing information:

- Update timer: when this timer expires, a RIP device immediately sends an Update packet.
- Age timer: if a RIP device does not receive any Update packet of a route from a neighbor within the Age timer, the RIP device considers this route unreachable.
- Garbage-collect timer: if a RIP device does not receive any Update packet of an unreachable route within the Garbage-collect timer, the device deletes this route from its RIP routing table.

The relationship between RIP routing and timer:

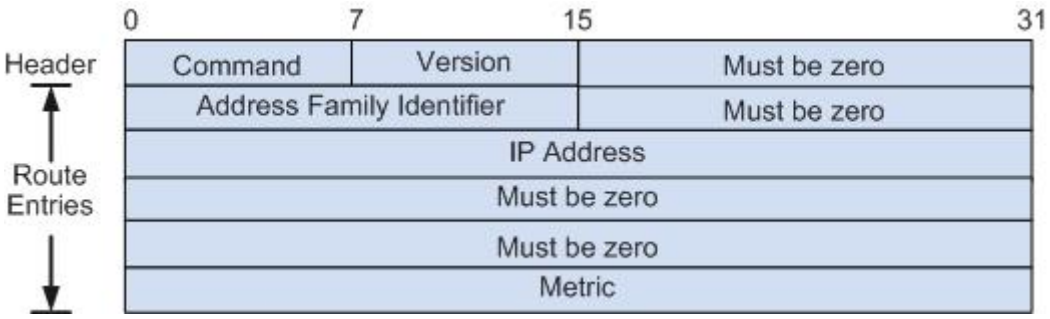
- The interval for sending Update packets is determined by the Update timer, which is 30 seconds by default.
- Each routing entry has two timers: Age timer and Garbage-collect timer. When a RIP device adds a learned route to the local RIP routing table, the Age timer starts for the route. If the RIP device does not receive an Update packet from the neighbor within the aging time, the RIP device sets the Cost value of the route to 16 (unreachable) and starts the Garbage-collect timer. If the RIP device still does not receive an Update packet within the Garbage-collect timer, the RIP device deletes the route from the RIP routing table.

16.2.2 RIP-2 Enhanced Features

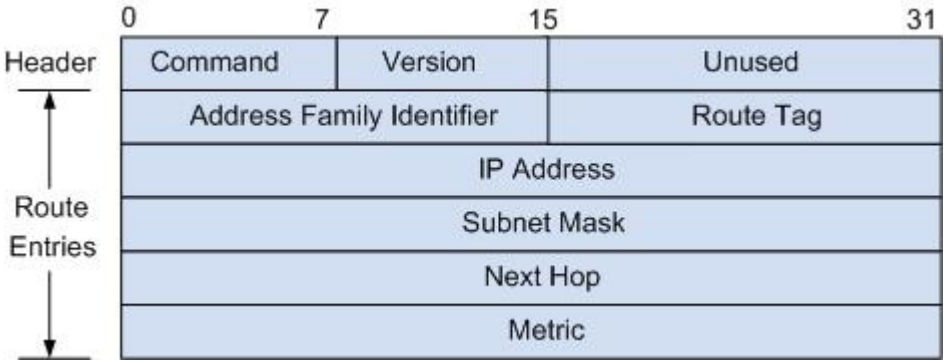
RIP includes two versions, RIP-1 and RIP-2. Rip-2 extends RIP-1.

16.2.2.1 Comparison Between RIP-1 and RIP-2

Rip-1 (also known as RIP Version1) is the Classful Routing Protocol, which only supports the broadcasting of Protocol packets. The format of packets is shown in the figure below. Because RIP-1 packets do not carry any mask information, they identify only the routes of natural network segments (such as Class A, Class B, and so on). Therefore, RIP-1 does not support route summarization or discontiguous subnets.



Rip-2 (also known as RIP version2) is a Classless Routing Protocol. The message format is shown in the figure below.



RIP-2 has the following advantages over RIP-1:

- Supports route tag and can flexibly control routes based on the tag in routing policies.
- Supports mask information and can therefore support route summarization and Classless Inter-Domain Routing (CIDR).
- Supports a next-hop address and can select the optimal next-hop address on broadcast networks.
- Supports the sending of update packets in multicast mode. Only RIP-2 devices can receive RIP-2 packets, saving resources.
- Provides packet authentication to enhance security.

16.2.2.2 RIP-2 Route Summarization

When different subnet routes within the same natural network segment are transmitted to other network segments, these routes are summarized into one route of the same network segment. This process is called route summarization. RIP-2 route summarization improves scalability and efficiency and reduces the size of the routing table for large networks.

RIP-2 route summarization is classified into two types:

- RIP process-based classful summarization

  Summarized routes are advertised using natural masks. For example, route 10.1.1.0/24 (metric=2) and route 10.1.2.0/24 (metric=3) are summarized as one route 10.0.0.0/8 (metric=2) in the natural network segment. RIP-2 supports classful summarization to obtain the optimal metric.
- Interface-based summarization

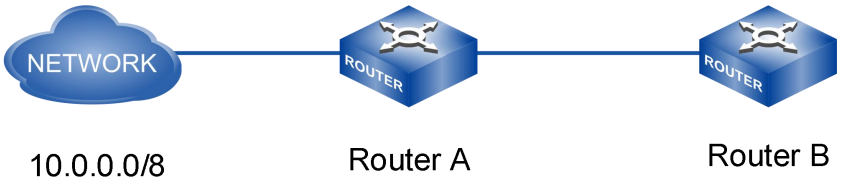
  You can specify a summarized address. For example, configure route 10.1.0.0/16 (metric=2) on an interface as the summarized route of route 10.1.1.0/24 (metric=2) and route 10.1.2.0/24 (metric=3).

16.2.3 Split Horizon and Poison Reverse

16.2.3.1 Horizon

Split horizon prevents a route learned by RIP on an interface from being sent to neighbors through this interface. It can not only reduce bandwidth consumption but also prevent routing loops.

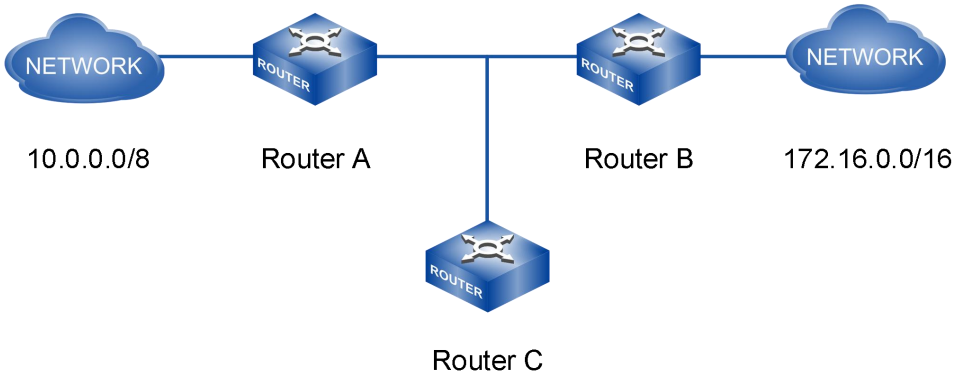
Horizontal segmentation is implemented differently in different networks, which can be divided into horizontal segmentation according to interfaces and horizontal segmentation according to neighbors. In broadcast networks, P2P and P2MP networks, horizontal segmentation is carried out according to interfaces, as shown in the figure below.



RouterA sends routing information destined for 10.0.0.0/8 to RouterB. If split horizon is not configured, RouterB sends the route learned from RouterA back to RouterA. Therefore, RouterA learns two routes destined for 10.0.0.0/8: a direct route with hop count 0 and a route with the next hop RouterB and hop count 2.

However, only the direct route in the RIP routing table of RouterA is active. If the route from RouterA to network 10.0.0.0 becomes unreachable, RouterB does not receive the route unreachable message immediately and continues to notify RouterA that network 10.0.0.0/8 is reachable. Therefore, RouterA receives incorrect routing information and considers that network 10.0.0.0/8 is reachable through RouterB, and RouterB considers that network 10.0.0.0/8 is reachable through RouterA, resulting in a routing loop. Using the split horizon function, RouterB does not send the route destined for 10.0.0.0/8 back to RouterA, avoiding a routing loop.

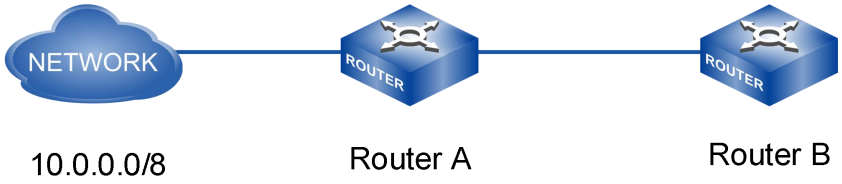
On a Non-Broadcast Multiple Access (NBMA) network, an interface connects to multiple neighbors; therefore, neighbor-based split horizon is used, as shown in Figure 2. Routes are advertised in unicast mode and the routes received by an interface are differentiated by neighbors. The route learned from a neighbor will not be sent back through the same interface.



As shown above, after the NBMA network is configured to split horizontally, RouterA sends the 172.16.0.0/16 route learned from RouterB to RouterC, but does not send it back to RouterB.

16.2.3.2 Poison Reverse

Poison reverse enables RIP to set the cost of the route learned through a local interface from a neighbor to 16 (unreachable) and then send this route through the same interface back to the neighbor. This function deletes invalid routes from the neighbor's routing table and prevents routing loops.



As shown in the figure above, after the configuration toxicity reverses, RouterB sends a message to RouterA that the route from RouterA is not reachable (setting the routing overhead to 16) so that RouterA no longer learns the route from RouterB and thus avoids the routing loop.

16.3 Configure RIP

16.3.1 Start RIP Process

【Command】

```
router rip
no router rip
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

router rip: this command is used to start the RIP process and enter RIP process mode.

Starting the RIP process is a prerequisite for all RIP configurations. If RIP related commands are configured in the interface view before starting RIP, these configurations will take effect only after RIP is started.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
```

16.3.2 Enable RIP in the Specified Network Segment

【Command】

network <A.B.C.D/M >

no network <A.B.C.D/M >

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D/M: network address and mask.

【Description】

network: this command is used to enable RIP on the specified network segment interface.

RIP function on the interface is disabled by default.

RIP runs only on interfaces on specified segments of the network. For interfaces that are not on the specified network segment, RIP neither receives nor sends its interface routes out. Therefore, RIP must specify its working network segment after starting.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch (config)#router rip
Switch (config- router)#network 10.11.20.0/24
```

16.3.3 Configure IP Address of RIP Neighbor in NBMA Network

【Command】

```
neighbor <A.B.C.D >
no neighbor <A.B.C.D >
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: neighbor interface address.

【Description】

neighbor: this command is used to configure the IP address of the RIP neighbor in the NBMA (non-broadcast multi-access) network and to send the update message to the opposite end as unicast rather than as normal multicast or broadcast.

no neighbor: this command is used to cancel the specified neighbor IP address.

By default, RIP does not send update message to any specified address.

Typically, RIP sends packets using broadcast or multicast addresses. If RIP is running on a link that does not support broadcast or group broadcasts, the NEIGHBORS of RIP must be manually assigned to each other at both ends of the link so that the message is sent unicast to the other end.

Notice

This command is not recommended when a RIP neighbor is directly connected to the current device, as it may cause the opposite side to receive both multicast (or broadcast) and unicast message with the same routing information.

【Instance】

```
Switch>enable
Switch#configure terminal
Switch(config)#router rip
Switch(config-router)#neighbor 10.11.20.1
```

16.3.4 Add Static RIP Route

【Command】

```
route <A.B.C.D /M >
no route <A.B.C.D /M>
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D /M:IP address and prefix.

【Description】

route: this command is used to add a static RIP route. This command is primarily for debugging purposes. The route configured by this command does not appear in the core routing table, but the route exists in the RIP routing database.

【Instance】

```
Switch>enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#route 10.11.20.1/16
```

16.3.5 Add Default Routing to RIP Routing Database

【Command】

```
default-information originate
no default-information originate
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

default-information originate: this command inserts the default route with a destination address of 0.0.0.0 into the RIP routing database and notifies the route like any other route.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#default-information originate
```

16.3.6 Default Route Metric

【Command】

default-metric <METRIC>
no default-metric

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

<METRIC> : sets the default metric when routing is introduced. <value> values range from 1 -16.

【Description】

default-metric: the command is used to set the default routing weight used to introduce routes from other routing protocols into the RIP route. When redistribute command is used to introduce a route for another protocol, if no specific route weight is specified, the default route weight specified by default-metric is introduced.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
```

```
Switch(config-router)#default-metric 10
```

16.3.7 RIP Route Management Distance

【Command】

```
distance <NUMBER> [ <A.B.C.D/M> ] [ <ACCESS-LIST-NAME> ]
no distance <NUMBER> [ <A.B.C.D/M> ] [ <ACCESS-LIST-NAME > ]
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

<NUMBER> : specifies the value of the distance, ranging from 1 to 255.

<A.B.C.D/ MB > : specifies the network prefix and prefix length.

< ACCESS-LIST-NAME > : specifies the name of the access list to be applied.

【Description】

distance: this command is used to set the RIP routing management distance.

no distance: this command restores the default value.

Administrative distance is used to select routes when there are routes from two different routing protocols reach the same destination. The smaller the management distance value of the routing protocol, the more reliable the routing obtained by the protocol.

By default, RIP administrative distance is 120.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#distance 8 10.0.0.0/8 mylist
```

16.3.8 Access List Route Filtering

【Command】

```
distribute-list ( <ACCESS-LIST-NAME> | prefix
<PREFIX-LIST-NAME> ) ( in | out ) [ <IFNAME> ]
no distribute-list ( <ACCESS-LIST-NAME> | prefix
<PREFIX-LIST-NAME> ) ( in | out ) [ <IFNAME> ]
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

< ACCESS-LIST-NAME > : the access list number or name of the application.

<PREFIX-LIST-NAME> : the name of the LIST of prefixes to apply.

<IFNAME>: specifies the interface name to which routing filtering is applied.

【Description】

distribute-list: the command is used to filter routing update message sent and received using an access list or prefix list. The no operation of this command is used to disable route filtering.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#distribute-list prefix myfilter in vlanif1
```

16.3.9 Other Routing Protocols Route Import

【Command】

```
redistribute { connected | static | ospf | bgp} [metric <VALUE> ]
[route-map <WORD>]
no redistribute { connected| static| ospf|bgp} [metric <VALUE>]
[route-map <WORD>]
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

connected：connected route

static: static route;

ospf：OSPF route.

bgp：BGP route.

<VALUE> : the measure assigned to the introduced route, with a value range of 0-16.

<WORD>：a pointer to the route map used to introduce routes.

【Description】

redistribute: the command introduces routes learned from other routing protocols into BGP.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#redistribute bgp metric 12
```

16.3.10 Block RIP Broadcast

【Command】

```
passive-interface <IFNAME>
no passive-interface <IFNAME>
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

<IFNAME>: interface name.

【Description】

passive-interface: the command is used to block RIP broadcast on the specified interface, so RIP packets can only be sent to the interface configured with neighbor. The no operation of this command is to disable this function.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#passive-interface vlanif2
```

16.3.11 Time of RIP Timer

【Command】

```
timers basic <UPDATE> <INVALID> <GARBAGE>
no timers basic
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

<UPDATE> : time interval for sending update message, the unit is second, value range: 5-2147483647.

<invalid> : the time period in which RIP route is declared invalid, the unit is second, and the value range is 5-2147483647.

<GARBAGE> : the time period that can still exist in the routing table after declaring a route invalid, the unit is second,the value range is 5-2147483647.

【Description】

timers basic: the command sets the RIP timer update, timeout, and garbage collection time. The no operation of this command is to restore to the default values of the parameters.

By default, <update> defaults to 30; Invalid > defaults to 180; <garbage> defaults to 120. The system broadcasts RIP update message every 30 seconds. When the update message of a route cannot be received after 180 seconds, the route is considered invalid. But the route can also exist in the routing table for 120 seconds, after which the routing table can delete it.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#timers basic 20 80 60
```

16.3.12 RIP Version

【Command】

```
version {1 | 2}
no version
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

1 is RIP version 1; 2 is RIP version 2.

【Description】

Version 1: indicates that each interface only sends/receives RIP-1 datagrams.

Version 2: indicates that each interface sends/receives RIP-2 datagrams only.

By default, RIP version 2 is sent and RIP-2 datagrams are received.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#version 1
```

16.3.13 Maximum Number of RIP Route

【Command】

maximum-prefix <MAXIMUM-PREFIX>[<THRESHOLD>]

no maximum-prefixd

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

<MAXIMUM-PREFIX> : the maximum number of RIP routes allowed, with a value range of 1-500;

<THRESHOLD> : when the percentage of the maximum number of routes exceeds, a warning will be generated by <threshold>. The value range is 1-100, and the default value is 75.

【Description】

maximum-prefix: the command is used to configure the maximum number of RIP routes in the routing table, and the no command removes the restriction on the number of routes.

The maximum number of RIP routes only limits the routes learned through RIP, excluding connected and introduced routes and RIP static routes configured with the route command. The comparison is based on the number of routes marked R in the

show IP route database command. And also is based on the RIP route number shown by the show IP route statistics command.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#maximum-prefix 150
```

16.3.14 RIP Routing Measures Offset

【Command】

```
offset-list < ACCESS-LIST-NAME> {in | out} <METRIC > [<IFNAME>]
no offset-list < ACCESS-LIST-NAME> {in | out} <METRIC > [<IFNAME>]
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

<ACCESS-LIST-NAME> : the name of the list of prefixes to be applied.

<METRIC> : the additional offset, and the value range is 0-16.

<IFNAME> : the specific interface name.

【Description】

offset-list: this command is used to configure the metric of the route learned through RIP plus an offset.

By default, RIP message is not validated.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#offset-list 1 in 5 vlanif1
```

16.3.15 RIP Route UDP to Receive Cache Size

【Command】

```
recv-buffer-size <SIZE>
no recv-buffer-size
```

【View】

RIP View

【Default Level】

2: Configuration level

【Parameter】

<SIZE> : buffer size in bytes, value range 8192-2147483647.

【Description】

recv-buffer-size: UDP receive buffer size of the command RIP; No operation restores to system defaults.

By default, it is 8192 bytes.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router rip
Switch(config- router)#recv-buffer-size 10240
```

16.3.16 RIP Message Authentication Mode

【Command】

```
ip rip authentication mode {text|md5}
no ip rip authentication mode [text|md5]
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

text: means text authentication.

md5: means MD5 authentication.

【Description】

ip rip authentication mode: the command sets the used type of authentication; the no operation of this command is to restore the default authentication type, that is text authentication.

By default, the relative interfaces do not configure passwords and keys.

RIP-1 does not support validation, and RIP-2 supports two types of authentication: text authentication (Simple authentication) and datagram authentication (MD5

authentication). This command needs to be used in combination with ip rip authentication key-chain or ip rip authentication string.Configuration alone does not perform authentication processing.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip rip authentication mode md5
```

16.3.17 RIP Message Authentication Key Chain

【Command】

```
ip rip authentication key-chain <NAME-OF-CHAIN>
no ip rip authentication key-chain [<NAME-OF-CHAIN>]
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

<NAME-OF-CHAIN> : the name of the key chain used, the string can contain spaces, the input ends with enter key, and the string length should not exceed 256.

【Description】

ip rip authentication key-chain: the command is used to enable RIPV2 authentication on an interface and to configure the used key chain. The no operation of this command is used to cancel authentication.

If authentication mode is configured only and the key chain or password used by the interface is not configured, authentication will not work at all. If mode is not set before this command is configured, it will be set to plaintext authentication. The no operation of this command will cancel authentication, it does not mean that mode will be set to the non-authenticated type, but authentication will not be processed when sending or receiving packets. The ip rip authentication key-chain my key command can be entered, which means the key chain name is my key, a total of 6 characters.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
```

```
Switch(config- vlanif1)#ip rip authentication key-chain my key
```

16.3.18 RIP Message Authentication Password

【Command】

```
ip rip authentication string <TEXT>
no ip rip authentication string
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

<TEXT> : the password used for authentication, with a length of 1-16 characters. Password can include space and end it with enter key.

【Description】

ip rip authentication string: the command is used to set the password used for RIP authentication. The no operation of this command is used to cancel authentication.

The ip rip authentication key chain command cannot be configured if this command is configured. The key id value is needed when using MD5 authentication, which is equivalent to 1 if use this command to configure the command. If mode is not set before this command is configured, it will be set to plaintext authentication. The no operation of this command will cancel authentication, it does not mean that mode will be set to the non-authenticated type, but authentication will not be processed when sending or receiving packets. The ip rip authentication string aaa aaa command can be entered, which means the key is aaa aaa, a total of 7 characters.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip rip authentication string guest
```

16.3.19 Receive RIP Message Enablement

【Command】

```
ip rip receive-packet
```

no ip rip receive-packet

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip rip receive-packet: this command is used to set whether the interface can receive RIP packets; The no operation of this command means cannot receive RIP message.

By default, the interface can receive RIP message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip rip receive-packet
```

16.3.20 Accept Message of Specified RIP Version

【Command】

ip rip receive version { 1 | 2 | 1 2 }
no ip rip receive version [1 | 2 | 1 2]

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

1 and 2 represent RIP version 1 and RIP version 2, respectively. 1 and 2 represent RIP version 1 and 2.

【Description】

ip rip receive-packet: the command is used to set the version information of the RIP packet received by the interface. RIP version 2 is received by default; The no operation of this command restores to the value set by the version command.

By default, the interface receives RIP message Version 2.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip rip receive version 1 2
```

16.3.21 Send RIP Message Enablement

【Command】

```
ip rip send-packet
no ip rip send-packet
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip rip send-packet: the command is used to set whether the interface can send RIP message; The no operation of this command means that RIP message cannot be sent.

By default, the interface can send RIP message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip rip send -packet
```

16.3.22 Send the Message of the Specified RIP Version

【Command】

```
ip rip send version { 1 | 2 | 1 2 }
no ip rip send version [ 1 | 2 | 1 2 ]
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

1 and 2 represent RIP version 1 and RIP version 2, respectively. 1 and 2 represent RIP version 1 and 2.

【Description】

ip rip send-packet: the command is used to set the version information of the RIP packet sent by the interface. RIP version 2 is sent by default; The no operation of this command restores to the value set by the version command.

By default, the version that the interface sends RIP message is 2.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip rip send version 1 2
```

16.3.23 RIP Horizontal Split Enablement

【Command】

```
ip rip split-horizon [ poisoned ]
no ip rip split-horizon
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

[poisoned] : it means the configuration with reverse poison horizontal segmentation.

【Description】

ip rip split-horizon: the command is used to enable horizontal split. The no operation of this command disables horizontal split.

Enables horizontal segmentation without reverse poisoning by default.

Horizontal segmentation is used to prevent Routing Loops, which prevent the interface of the device from broadcasting routes learned from itself. In general, horizontal segmentation is necessary to prevent routing loops, so it is not recommended to disable it. When it is necessary to disable horizontal segmentation

for special reasons, such as ensuring proper execution of the protocol, be sure to confirm whether it is necessary.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip rip split-horizon poisoned
```

16.3.24 Display Routing Information learned by RIP

【Command】

show ip rip

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show ip rip: the command is used to display the routing information learned by rip protocol.

Notice

The show ip route rip command is used to display the route information that learned by rip protocol and show in the routing table.

【Instance】

```
SwitchA> enable
SwitchA#configure terminal
SwitchA(config)#interface vlanif1
SwitchA(config- vlanif1)#ip address 192.168.1.1/24
SwitchA(config- vlanif1)#exit
SwitchA(config)#vlan database
SwitchA(config-vlan)#vlan 2
SwitchA(config-vlan)#exit
SwitchA(config)#ip interface vlan 2
```

```
SwitchA(config)#interface vlanif2
SwitchA(config- vlanif2)#ip address 192.168.2.2/24
SwitchA(config- vlanif2)#exit
SwitchA(config)#router rip
SwitchA(config-router)#network 192.168.1.0/24
SwitchA(config-router)#network 192.168.2.0/24

SwitchB> enable
SwitchB#configure terminal
SwitchB(config)#vlan database
SwitchB(config-vlan)#vlan 2
SwitchB(config-vlan)#exit
SwitchB(config)#ip interface vlan 2
SwitchB(config)#interface vlanif2
SwitchB(config- vlanif2)#ip address 192.168.2.1/24
SwitchB(config- vlanif2)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#network 192.168.2.0/24
SwitchB(config-router)#end
Switch#show ip rip

Codes: R - RIP, Rc - RIP connected, Rs - RIP static, K - Kernel,
      C - Connected, S - Static, O - OSPF, I - IS-IS, B - BGP

Network          Next Hop        Metric From           If     Time
R   192.168.1.0/24       192.168.2.2               2  192.168.2.2
vlanif2 02:41
Rc 192.168.2.0/24                        1              vlanif2

Switch#show ip route rip
R      192.168.1.0/24 [120/2] via 192.168.2.2, vlanif2, 00:00:32
```

16.3.25 Display the Routing Information in the RIP Routing Information Base

【Command】

show ip rip database

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show ip rip database: the command is used to display routing information in the rip routing information database.

【Instance】

```
SwitchA> enable
SwitchA#configure terminal
SwitchA(config)#interface vlanif1
SwitchA(config- vlanif1)#ip address 192.168.1.1/24
SwitchA(config- vlanif1)#exit
SwitchA(config)#vlan database
SwitchA(config-vlan)#vlan 2
SwitchA(config-vlan)#exit
SwitchA(config)#ip interface vlan 2
SwitchA(config)#interface vlanif2
SwitchA(config- vlanif2)#ip address 192.168.2.2/24
SwitchA(config- vlanif2)#exit
SwitchA(config)#router rip
SwitchA(config-router)#network 192.168.1.0/24
SwitchA(config-router)#network 192.168.2.0/24

SwitchB> enable
SwitchB#configure terminal
SwitchB(config)#vlan database
SwitchB(config-vlan)#vlan 2
SwitchB(config-vlan)#exit
SwitchB(config)#ip interface vlan 2
SwitchB(config)#interface vlanif2
SwitchB(config- vlanif2)#ip address 192.168.2.1/24
SwitchB(config- vlanif2)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#network 192.168.2.0/24
SwitchB(config-router)#end
Switch#show ip rip database

Codes: R - RIP, Rc - RIP connected, Rs - RIP static, K - Kernel,
       C - Connected, S - Static, O - OSPF, I - IS-IS, B - BGP
```

```
Network      Next Hop      Metric From       If              Time
R  192.168.1.0/24   192.168.2.2   2  192.168.2.2  vlanif2   02:57
Rc 192.168.2.0/24                      1              vlanif2
```

16.3.26 Display RIP Interface Information

【Command】

show ip rip interface <IFNAME>

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show ip rip interface: the command is used to display rip interface information.

【Instance】

```
SwitchA> enable
SwitchA#configure terminal
SwitchA(config)#interface vlanif1
SwitchA(config- vlanif1)#ip address 192.168.1.1/24
SwitchA(config- vlanif1)#exit
SwitchA(config)#vlan database
SwitchA(config-vlan)#vlan 2
SwitchA(config-vlan)#exit
SwitchA(config)#ip interface vlan 2
SwitchA(config)#interface vlanif2
SwitchA(config- vlanif2)#ip address 192.168.2.2/24
SwitchA(config- vlanif2)#exit
SwitchA(config)#router rip
SwitchA(config-router)#network 192.168.1.0/24
SwitchA(config-router)#network 192.168.2.0/24

SwitchB> enable
SwitchB#configure terminal
SwitchB(config)#vlan database
SwitchB(config-vlan)#vlan 2
```

```
SwitchB(config-vlan)#exit
SwitchB(config)#ip interface vlan 2
SwitchB(config)#interface vlanif2
SwitchB(config- vlanif2)#ip address 192.168.2.1/24
SwitchB(config- vlanif2)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#network 192.168.2.0/24
SwitchB(config-router)#end
Switch#show ip rip interface vlanif2
vlanif2 is up, line protocol is up
  Routing Protocol: RIP
    Receive RIP packets
    Send RIP packets
    Passive interface: Disabled
    Split horizon: Enabled
    IP interface address:
      192.168.2.1/24
```

17 OSPF Configuration

17.1 Overview

The Open Shortest Path First (OSPF) protocol is link-state Interior Gateway Protocol (IGP) developed by the Internet Engineering Task Force (IETF).

OSPF Version 2 as defined in RFC 2328 is designed for IPv4. OSPF Version 3 as defined in RFC 2740 is designed for IPv6. Unless otherwise stated, OSPF in this document refers to OSPF Version 2.

- Dividing an Autonomous System (AS) into one or more logical areas
- Advertising routes by sending Link State Advertisements (LSAs)
- Exchanging OSPF packets between devices in an OSPF area to synchronize routing information
- Encapsulating OSPF packets into IP packets and then sending the packets in unicast or multicast mode

Prior to the emergence of OSPF, the Routing Information Protocol (RIP) was the most widely used IGP.

RIP is a distance-vector routing protocol. Due to RIP's slow convergence, routing loops, and poor scalability, OSPF is now the most widely accepted and used IGP.

OSPF, as a link-state based protocol, can solve many problems faced by RIP. In addition, OSPF has the following advantages:

- Multicast packet transmission to reduce load on the switches that are not running OSPF
- Classless Inter-Domain Routing (CIDR)
- Load balancing among equal-cost routes
- Packet authentication

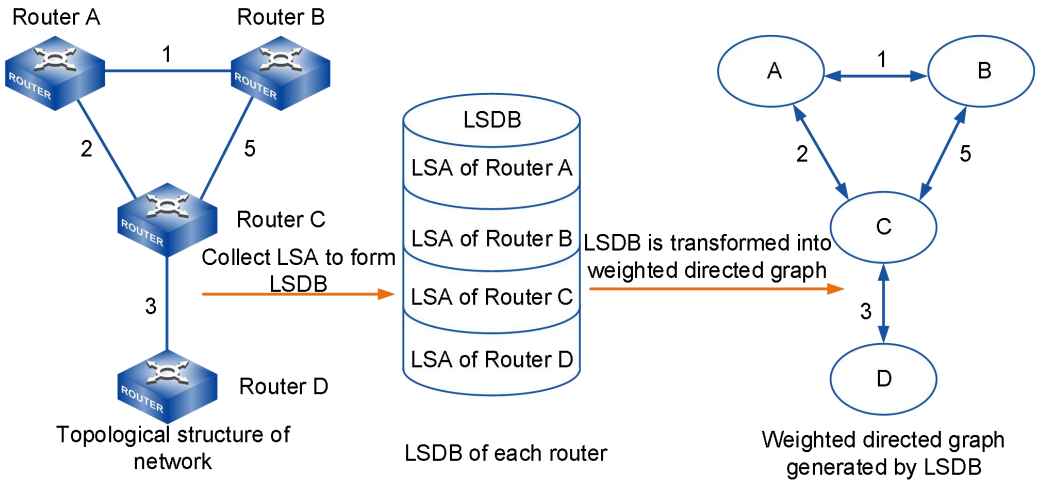
17.2 Principle Description

17.2.1 OSPF Characteristics

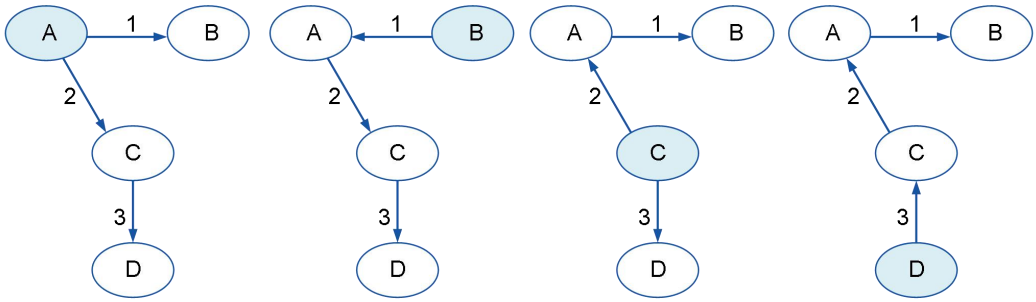
In an OSPF network, each router generates a link-state advertisement (LSA) based on its surrounding network topology and transmits this LSA in an update packet to other routers in the network.

RIP interacts with routing. Unlike RIP, OSPF interacts with link state information. That is, in RIP, routers select routes based on routing information of neighbors, without checking whether the information transmitted by neighbors is correct; in OSPF, routers calculate routes by themselves and select routes based on LSAs.

Each router learns about the whole network topology based on its link state database (LSDB). As shown in the following figure, each router collects LSA sent by other routers, and all LSA are put together to form a link state database LSDB. LSA is a description of the network topology around the router, while LSDB is a description of the network topology of the whole autonomous system. A router transforms its LSDB into a weighted, directed graph, which reflects the topology of the entire AS. When the network topology is stable, all routers in the same area have the same graph.



A router calculates shortest paths to destinations using a shortest path first (SPF) algorithm instead of obtaining routing information through route advertisements. As shown in the following figure, each router uses SPF algorithm to calculate a shortest path tree with itself as the root according to the directed graph, which gives the route to each node in the autonomous system. This mechanism greatly improves routers' capabilities in independent route selection so that routers do not need to select routes based on route advertisement.



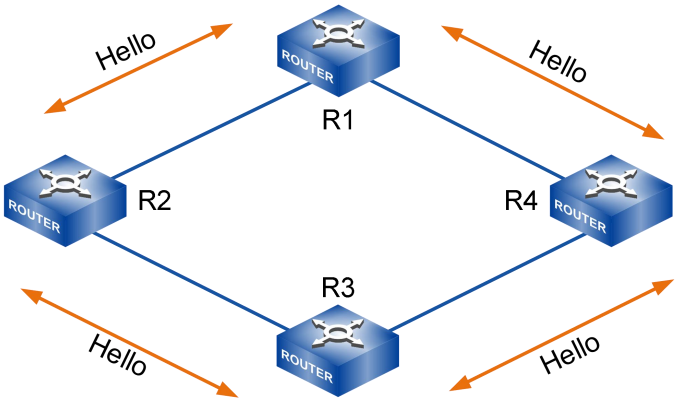
An LSDB ensures that a router learns about the whole network topology, and an SPF algorithm ensures that a router fast calculates shortest paths to destinations.

17.2.2 OSPF Mechanism

The OSPF mechanism includes the following steps:

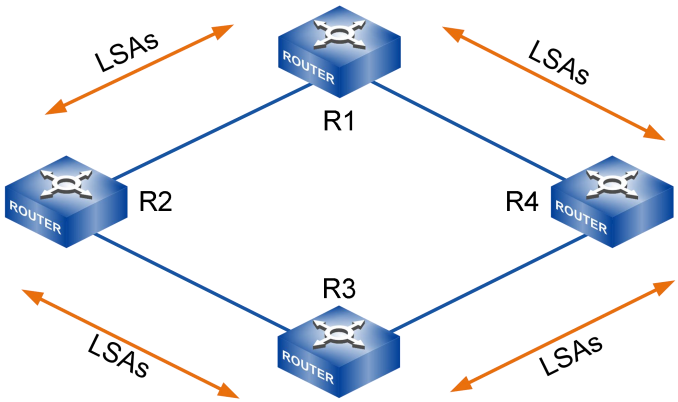
Step 1 Exchanging Hello packets to establish OSPF neighbor relationships

As shown in the following figure, after the router runs OSPF protocol, it will send Hello messages from all interfaces that start OSPF protocol. If two routers share a data link and successfully negotiate some parameters specified in their Hello packets, they establish an OSPF neighbor relationship.



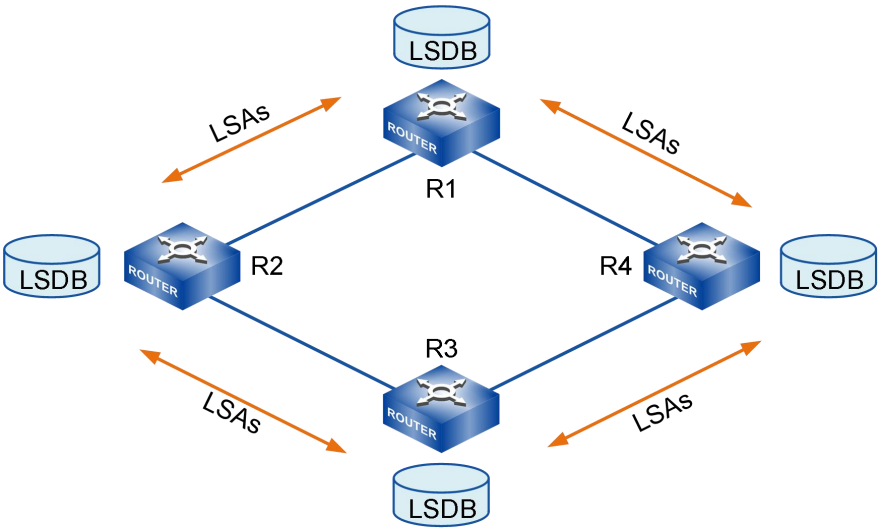
Step 2 Flooding LSAs to advertise link state information

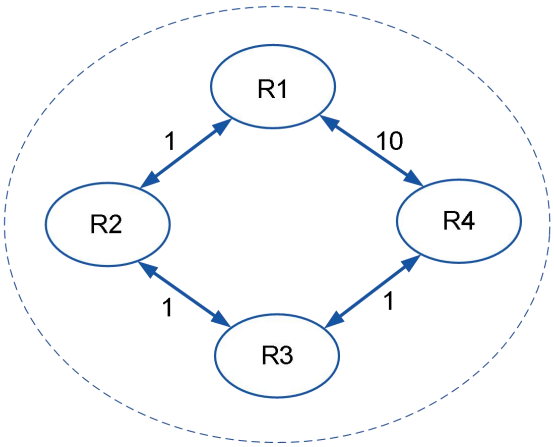
LSA can be exchanged between routers forming adjacency relationship, as shown in the following figure. LSAs describe information about a router, including all links, interfaces, neighbors, and link state. Routers exchange the link information to learn about the whole network topology. Because of link diversity, OSPF defines multiple LSA types.



Step 3 Routers form their LSDBs to create a weighted, directed graph.

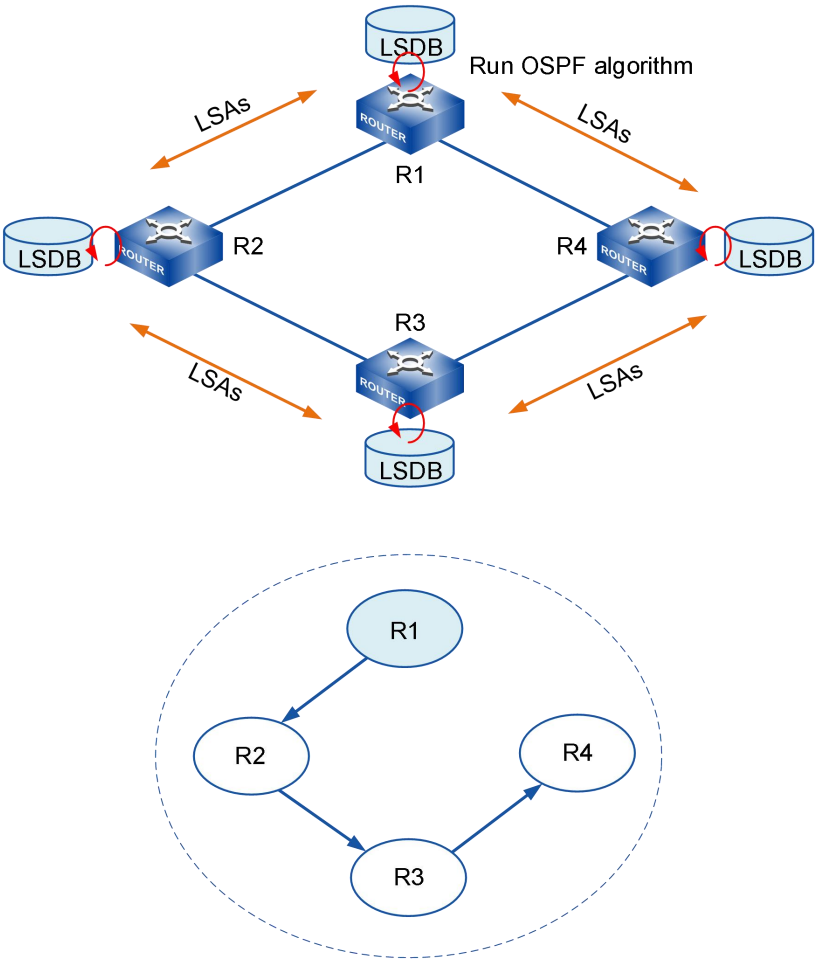
A router floods LSAs and records received LSAs in its LSDB. Eventually, all routers will form the same LSDB, as shown in the following figure. An LSA describes the surrounding network topology of a router, whereas an LSDB describes the network topology of the entire AS and is the summary of LSAs.





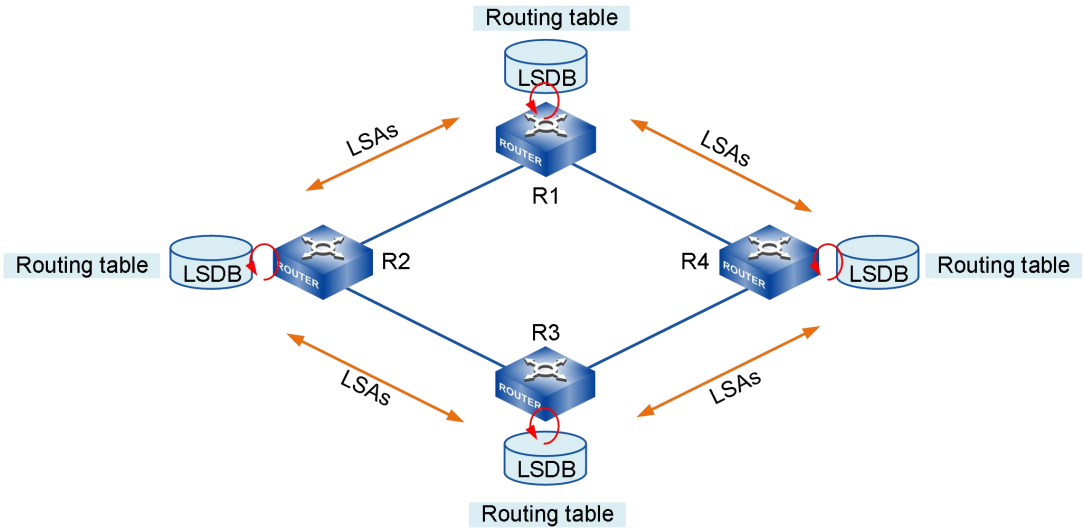
Step 4 Using an SPF algorithm to calculate and generate routes

As shown in the following figure, after LSDB synchronization is completed, each router will use SPF algorithm to calculate a loop-free topological graph to describe the shortest path (minimum path cost) it knows to reach each destination. The topology is the SPT, which shows the optimal paths to nodes in an AS.



Step 5 Maintaining and updating routing tables

After obtaining the shortest path tree according to SPF algorithm, each router loads the calculated shortest path into OSPF routing table to form a routing table entry guiding data forwarding, and updates it in real time, as shown in the following figure. Meanwhile, neighbors exchange Hello packets to maintain their neighbor relationships or adjacencies and periodically retransmit LSAs.



17.2.3 OSPF Packet Types

OSPF provides five types of packets:

- Hello packet

  Hello message is the most commonly used message, which is used to establish and maintain adjacency relationship. Interfaces with OSPF enabled send Hello messages periodically. These packets contain some timer values and information about the Designated Router (DR), Backup Designated Router (BDR), and known neighbors on the same network.

- Database Description (DD) packet

  During adjacency initialization, two routers send DD packets to negotiate their master/slave relationship. The DD packets do not contain an LSA header. When two routers exchange DD packets, one functions as the master and the other functions as the slave. The master defines a start sequence number and increases the sequence number by 1 each time it sends a DD packet. After the slave receives a DD packet, it uses the sequence number carried in the DD packet for acknowledgement.

  After an adjacency is established, routers use DD packets to describe their own LSDBs for LSDB synchronization. A DD packet contains the header of each LSA in an LSDB and is the summary of all LSAs. An LSA header uniquely identifies an LSA. The LSA header occupies only a small portion of the LSA, which reduces the amount of traffic transmitted between routers. A neighbor can use the LSA header to check whether it already has the LSA.

- Link State Request (LSR) packet

  After two routers have exchanged DD packets, they send LSR packets to request

each other's LSAs. The LSR packets contain the summaries of the requested LSAs.

- Link State Update (LSU) packet
  A router uses an LSU packet to transmit LSAs requested by its neighbors or to flood its own updated LSAs. The LSU packet contains a set of LSAs. To ensure reliable LSA flooding, the router uses an LSAck packet to acknowledge the LSAs contained in an LSU packet that is received from a neighbor. If an LSA fails to be acknowledged, the router retransmits the LSA to the neighbor.
- Link State Acknowledgement (LSAck) packet
  A router uses an LSAck packet to acknowledge the LSAs contained in a received LSU packet. One LSAck message can validate multiple LSAs.

17.2.4 OSPF Network Types

Four OSPF network types are classified based on link layer protocols.

- Broadcast
  Networks using Ethernet or Fiber Distributed Data Interface (FDDI) at the link layer are broadcast networks by default.
  On this type of networks:
  - Hello packets, LSU packets, and LSAck packets are sent in multicast mode. 224.0.0.5 is the IP multicast address reserved for OSPF devices. 224.0.0.6 is the IP multicast address reserved for OSPF DRs or BDRs.
  - DD and LSR packets are sent in unicast mode.
- Non-Broadcast Multi-Access (NBMA)
  Networks using frame relay (FR) or X.25 at the link layer are NBMA networks by default.
  Within NBMA networks, protocol packets such as Hello packets, DD packets, LSR packets, LSU packets, and LSAck packets are sent in unicast mode.
- Point-to-Multipoint (P2MP)
  No networks are P2MP networks by default, regardless of the link layer protocol used by the network. Point-to-multipoint changes must be forced by other network types. Typical practice is to change partial-meshed NBMA networks to P2MP networks.
  On this type of networks:
  - Hello packets are transmitted in multicast mode using the multicast address 224.0.0.5.
  - DD packets, LSR packets, LSU packets, and LSAck packets are sent in unicast mode.

- Point-to-Point (P2P)
  Networks using Point-to-Point Protocol (PPP), High-Level Data Link Control (HDLC), or Link Access Procedure Balanced (LAPB) at the link layer are P2P networks.
  Within P2P networks, protocol packets such as Hello packets, DD packets, LSR packets, LSU packets, and LSAck packets are sent in multicast mode using the multicast address 224.0.0.5.

17.2.5 DR/BDR Election

17.2.5.1 Router ID

Before learning the DR/BDR election process, you need to know the router ID. A router ID is a 32-bit integer, which uniquely identifies an OSPF router in an AS. Each OSPF router has a router ID. A router ID is in the same format as an IP address. To ensure OSPF stability in actual network deployment, it is recommended that the IP address of a loopback interface on a router be used as the router ID of this router.
A router ID can be manually configured or automatically selected by a router.
If no router ID is manually configured for a router, the router automatically selects an interface IP address as its router ID. The router ID selection rules are as follows:

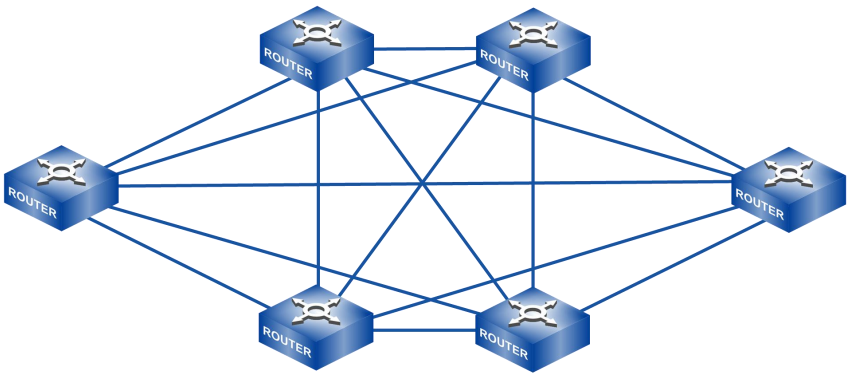
1 The router preferentially selects the largest IP address among loopback interface addresses as the router ID.
2 If no loopback interface is configured, the router selects the largest IP address among interface addresses as the router ID.

A switch can obtain a router ID again only after a router ID is reconfigured for the switch or an OSPF router ID is reconfigured and the OSPF process restarts.

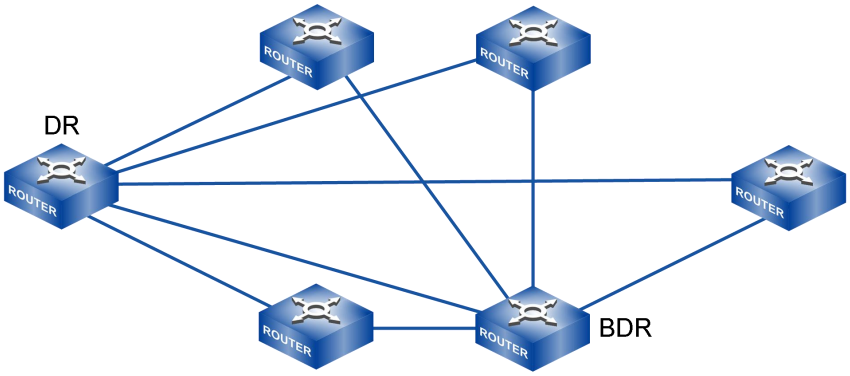
17.2.5.2 Reason for DR/BDR Election

On broadcast and non-broadcast multiple access (NBMA) networks, any two routers need to exchange routing information. On the network in the following figure, there are n routers that need to establish n x (n - 1)/2 adjacencies. This causes any router's routing changes to result in multiple passes, wasting bandwidth resources.
To solve this problem, the concept of DR is defined in OSPF. After a DR is elected, all the other routers send routing information only to the DR, and the DR broadcasts LSAs.

To prevent service interruption caused by DR re-election when the DR fails, a BDR is also elected during DR election. The routers excluding DR and BDR are called DR others. The DR others do not establish adjacencies or exchange any routing information with each other. This reduces the number of adjacencies established between routers on broadcast and NBMA networks.



Before electing DR/BDR



After electing DR/BDR

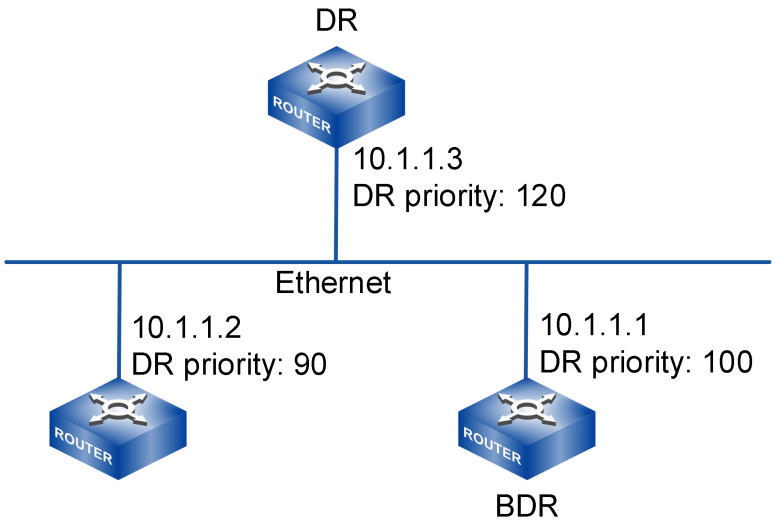
17.2.5.3 DR/BDR Election Principles

To ensure stable DR/BDR election on broadcast and NBMA networks, OSPF defines three election principles: election, non-preemption, and inheritance.

- Election Principle

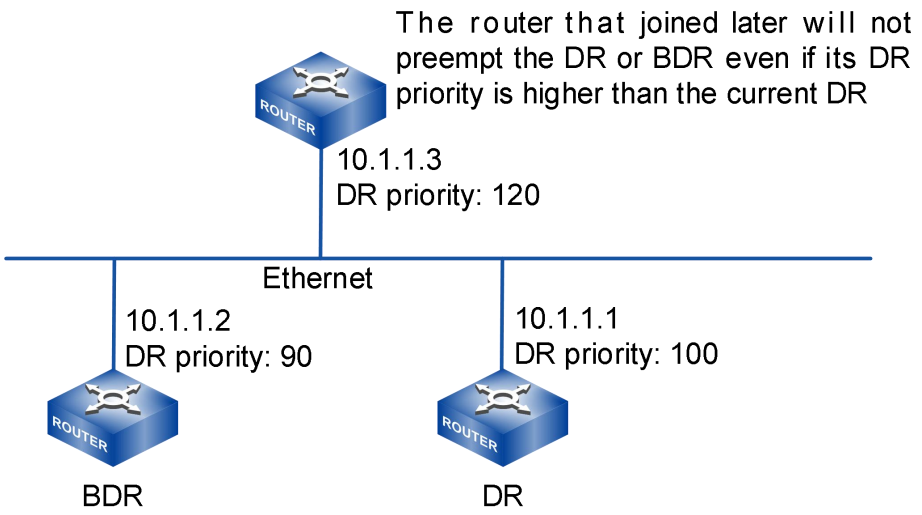
  The DR and BDR are not designated manually, but are elected by all routers on the local network segment. In the following figure, DR priorities of router interfaces determine whether the interfaces are qualified for DR/BDR election. Routers with DR priorities larger than 0 on the local network segment can be considered as candidates. The ballots in this election are Hello packets. Each router adds elected DR information to a Hello packet and sends the packet to other routers on the network segment. When two routers on the same network segment declare that they are the DR, the router with a higher DR priority is

elected as the DR. If the priority is equal, the one with larger Router ID wins. If a router has a priority of 0, it will not be elected DR or BDR.



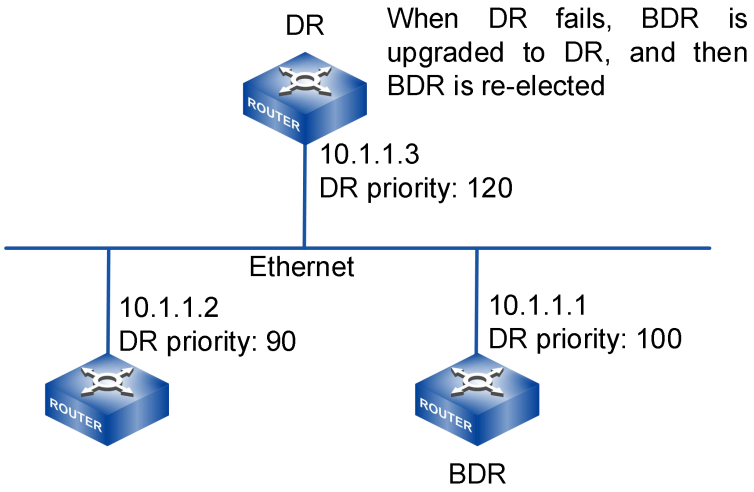
- Non-preemption Principle

  A router newly added to a network segment does not attempt in election and checks whether a DR exists on the network segment. In Figure 3, a DR exists on the network segment. As shown in the figure below, even if the DR priority of the newly added router is higher than that of the DR, the newly added router does not declare itself as the DR, and acknowledges the existing DR. Routers on a network segment only establish adjacencies with the DR and BDR. If the DR changes frequently, all routers on the network segment need to establish adjacencies with the new DR and BDR accordingly. As a result, a large number of OSPF packets are transmitted on the network segment within a short period of time, reducing the available bandwidth. The non-preemption principle improves network stability and saves the available network bandwidth. On a broadcast or NBMA network, the two routers that are qualified for DR election and start first become the DR and BDR.



- Inheritance Principle

  If the DR on a network segment fails, the BDR is elected as the DR, and the other routers compete to become the new BDR, as shown in Figure. This principle ensures the DR stability and prevents frequent DR elections. If the DR has a BDR, the BDR becomes the new DR when the DR fails. Because LSDBs of the DR and BDR are fully synchronized, the BDR immediately becomes the new DR when the DR fails and performs duties of the DR. Because adjacencies have been established, the time from role change to service switching is short. In addition, when the BDR becomes the new DR, a new BDR will be elected. Although the election process takes a relatively long time, route calculation is not affected.



17.2.5.4 DR/BDR Election Process

The DR/BDR election process on a broadcast or NBMA network is as follows:

3 After an interface goes Up, it sends a Hello packet and enters the Waiting state.

In the Waiting state, a wait timer is triggered. The wait timer value is the same as the dead timer value. The default wait timer value is 40 seconds and cannot be changed.

4 Before the wait timer is triggered, sent Hello packets do not contain DR or BDR information. In the Waiting state, if a received Hello packet contains DR and BDR information, the interface directly acknowledges the DR and BDR on the network, and does not trigger election. The interface directly exits the Waiting state and starts neighbor synchronization.

5 Assume that a DR and a BDR exist on the network. A router newly connected to the network acknowledges the existing DR and BDR regardless of how large its router ID or DR priority is.

6 If the DR fails and goes down, the BDR takes over the role of the DR and the other routers whose priority is greater than 0 compete to become the new BDR.

7 DR election rules are used to elect a DR only when routers with different router IDs or configured with different DR priorities are started at the same time. The election rules are that the device with the highest DR priority is elected as DR and the device with the second highest DR priority as BDR. A router with a DR priority of 0 can be a DR other only. If routers have the same DR priority, the router with the greatest router ID is elected as the DR, the router with the second greatest router ID becomes the BDR, and other routers are DR others.

17.2.6 OSPF State Machine

17.2.6.1 OSPF interface state machine

An OSPF device obtains link information from an interface and then establishes an adjacency with neighbors to exchange the link information. Before establishing adjacencies, these devices need to determine their roles to establish connections. The State field of OSPF interface information indicates the role of the OSPF device in the corresponding link.

The OSPF interface state machine has the following states:

- Down: This state is the initial state of an OSPF interface. If an OSPF interface is Down, it is unavailable and cannot receive and transmit traffic.
- Loopback: An OSPF interface that connects a device to the network is in Loopback state. A loopback interface cannot transmit data but can be advertised using Router-LSAs. Therefore, during a connectivity test, you can find the path to a loopback interface.
- Waiting: The device is determining the DR and BDR on the network. Before the

device participates in DR and BDR election, the wait timer starts on an OSPF interface. Before this timer expires, the Hello packets sent by this device do not contain DR and BDR information, and the device cannot be elected as the DR or BDR. This prevents the existing DR and BDR on the link from being changed unnecessarily and ensures stability. This state exists only on Non-Broadcast Multi-Access (NBMA) and broadcast networks.

- P-2-P: An OSPF interface is connected to a physical point-to-point network or virtual link. In this situation, the device establishes an adjacency with the device on the other end of the link. This state exists only on P2P and point-to-multipoint (P2MP) networks.
- DROther: The device is not elected as the DR or BDR, and another device that is connected to a broadcast or NBMA network is elected as the DR. This device will establish an adjacency with the DR and BDR.
- BDR: The device functions as the BDR on the network and will become the new DR when the existing DR fails. This device establishes adjacencies with all the other devices on the network.
- DR: The device functions as the DR on the network. This device establishes adjacencies with all the other devices on the network.

An OSPF interface alternates among the preceding states based on the InputEvent (IE). This forms an efficient interface state machine, as shown in the Figure.

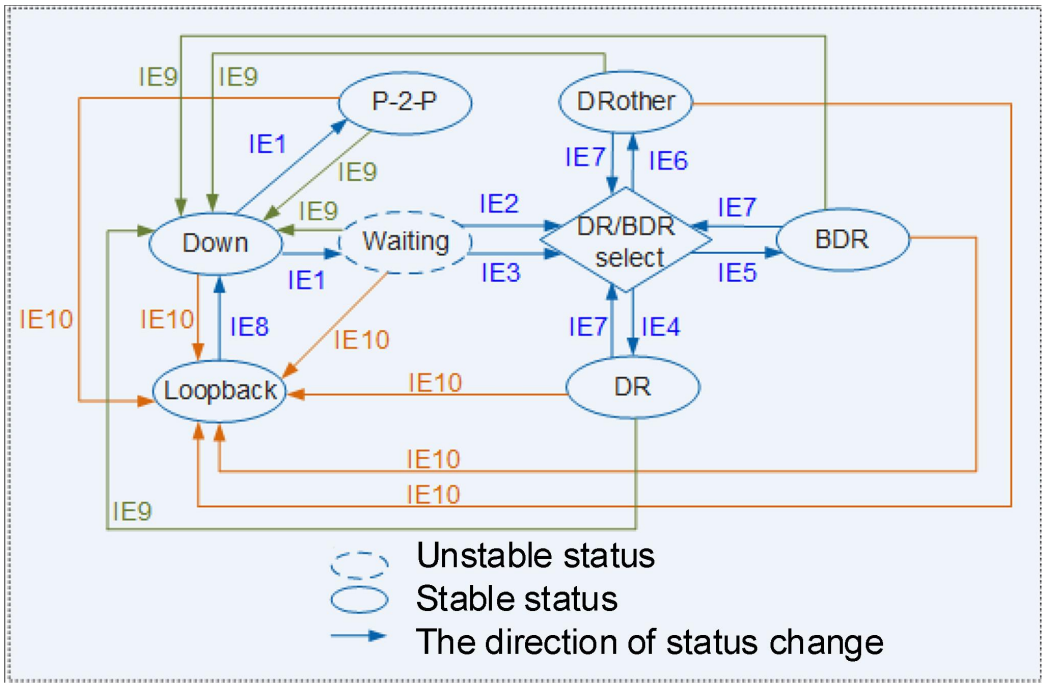


Table lists the InputEvents occurring during a neighbor state switchover. The InputEvent is abbreviated to IE in the above figure.

InputEvent	Desc
IE1	InterfaceUP: Lower-layer protocols indicate that an OSPF interface is

InputEvent	Desc
	available.
IE2	WaitTimer: The wait timer expires, indicating that the DR/BDR election waiting time ends.
IE3	BackupSeen: The device where an OSPF interface is located has detected whether there is a BDR on the network. This event occurs in either of the following situations: • The OSPF interface receives a Hello packet from a neighbor, and the neighbor declares itself as the BDR. • The OSPF interface receives a Hello packet from a neighbor, and the neighbor declares itself as the DR but no BDR is specified. The two situations indicate that neighbors have communicated with each other and the Waiting state ends.
IE4	The device is elected as the DR on the network.
IE5	The device is elected as the BDR on the network.
IE6	The device is not elected as the DR or BDR on the network.
IE7	NeighborChange: A neighbor relationship change event related to the interface occurs, indicating that the DR and BDR need to be elected again. The following neighbor relationship changes may result in DR/BDR re-election: • The device establishes bidirectional communication with a neighbor. • Bidirectional communication between the device and a neighbor is interrupted. • The device detects that the neighbor declares itself as the DR or BDR based on the Hello packet received from the neighbor. • The device detects that the neighbor declares itself not the DR or BDR based on the Hello packet sent from the neighbor. • The device detects that the DR priorities of all neighbors have changed based on the Hello packets sent by the neighbors.
IE8	UnLoopInd: The NMS or lower-layer protocols indicate that the interface is not in Loopback state.
IE9	InterfaceDown: Lower-layer protocols indicate that the interface is unavailable. Any state may change to the Down state after this event

InputEvent	Desc
	is triggered.
IE10	LoopInd: The NMS or lower-layer protocols indicate that the interface is in Loopback state. Any state may change to the Loopback state after this event is triggered.

17.2.6.2 OSPF neighbor state machine

On an OSPF network, neighbors establish adjacencies and exchange LSAs through neighbor state switchovers.

The State field in OSPF neighbor information indicates the neighbor status of an OSPF device.

The OSPF neighbor state machine has the following states:

- Down: This is the initial state of a neighbor session. This state occurs when a device does not receive any Hello packets from its neighbors within a dead interval. Only OSPF routers on an NBMA network send Hello packets to neighbors at each poll interval, including the neighbors in Down state.
- Attempt: This state is only used on NBMA networks where neighbors are manually configured. When the neighbor relationship is in Attempt state, an OSPF router sends a Hello packet to its manually configured neighbors at each Hello interval, trying to establish a neighbor relationship.
- Init: This state occurs after the router has received a Hello packet from its neighbor but has not established a two-way session. In this state, the neighbor does not receive any Hello packet from this router, and the neighbor list in the received Hello packet does not contain the router ID of the local router.
- 2-Way: Neighbors. This state occurs when a router and its neighbor receive Hello packets containing their own router IDs from each other and establish an OSPF neighbor relationship. If no adjacency needs to be established, the two neighbors remain in the 2-way state. If adjacencies need to be established, the neighbors enter the Exstart state. The DR and BDR are elected only when the neighbor state is the 2-way state or higher.
- ExStart: Negotiate master-slave relationship. This state occurs when the two neighbors start to negotiate the master/slave status and determine the sequence numbers of DD packets. Exstart is the first step in creating an adjacency.
- Exchange: Exchange DD messages. This state occurs when the two neighbors start to exchange DD packets. DD packets contain LSDB information.
- Loading: synchronizing LSDBs. The two devices send LSR packets to request LSAs from each other to synchronize their LSDBs.

- Full: Establish adjacency. This state occurs when the two neighbors establish an adjacency after their LSDB synchronization is completed.

Figure shows an OSPF neighbor state switchover.

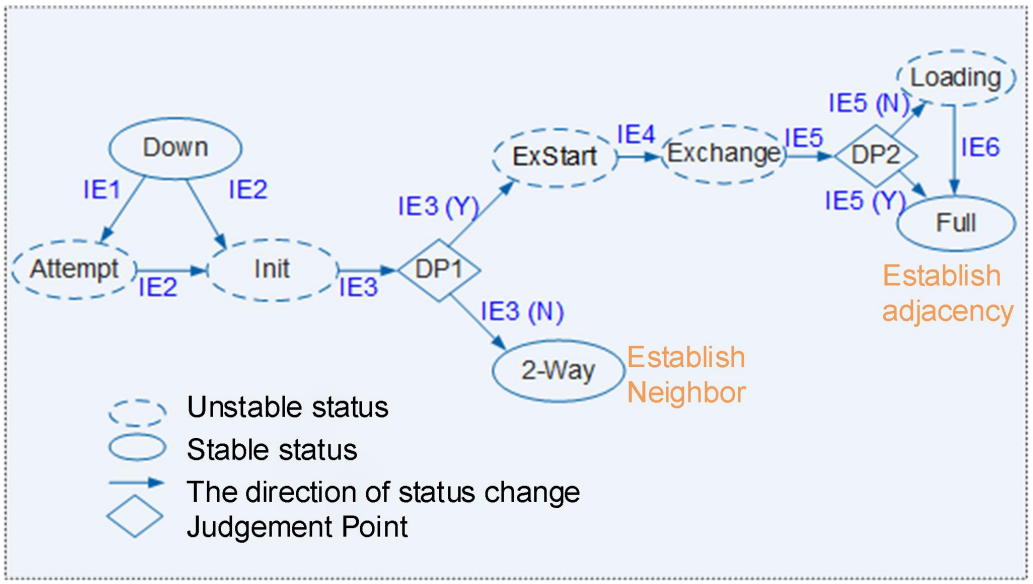


Table lists the InputEvents occurring during a neighbor state switchover. The InputEvent is abbreviated to IE in the above figure.

InputEvent	Desc
IE1	Start: A router sends a Hello packet to its neighbors at each Hello interval, trying to establish a neighbor relationship. This event exists only on an NBMA network.
IE2	HelloReceived: The router receives a Hello packet from a neighbor.
IE3	2-WayReceived: The router receives a Hello packet containing its router ID from a neighbor and establishes a two-way session with this neighbor. Then the router determines its neighbor state accordingly: • IE3(Y): If it needs to establish an adjacency with the neighbor, it changes its neighbor state to Exstart. • IE3(N): If it does not need to establish an adjacency with the neighbor, it changes its neighbor state to 2-way.
IE4	NegotiationDone: The two neighbors have negotiated their master/slave roles and exchanged sequence numbers of DD packets.
IE5	ExchangeDone: The two neighbors have exchanged their DD packets. Then the router determines its neighbor state accordingly:

InputEvent	Desc
	• IE5(Y): If the link state request list is empty, the router changes its neighbor state to Full, indicating that all link state data has been exchanged and neighbors have established adjacencies. • IE5(N): If the link state request list is not empty, the router changes its neighbor state to Loading and begins or continues sending LSR packets to its neighbors to request link state data that it does not have.
IE6	LoadingDone: The link state request list is empty.

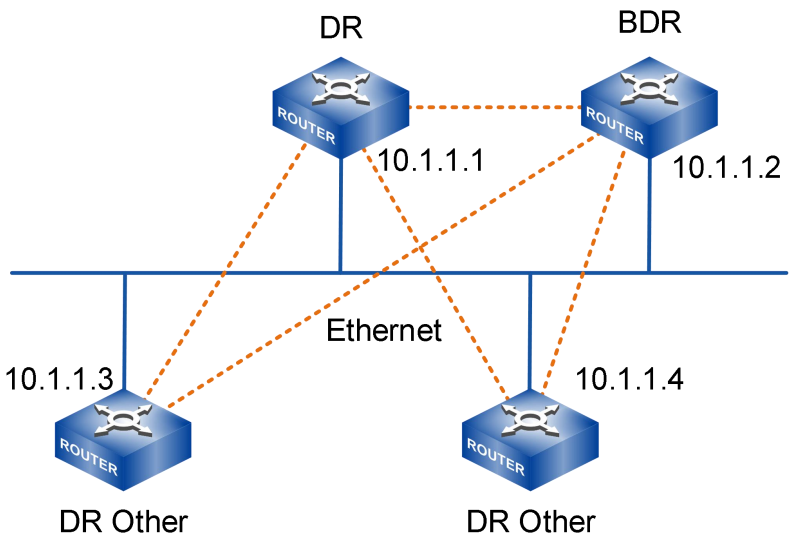
17.2.7 OSPF Adjacency Establishment

17.2.7.1 OSPF neighbor relationship and adjacency

After an OSPF device starts, it sends a Hello packet through its OSPF interface. After another OSPF device on the same network receives this Hello packet, it checks parameters defined in this packet, including the interval for sending Hello packets, interface addresses of the DR and BDR, and IP address mask. If these parameters are consistent in the Hello packets of both ends, the two devices establish an OSPF neighbor relationship and become neighbors.

OSPF adjacency relationships are above neighbor relationships. The two neighbors need to exchange DD packets and LSAs to establish an adjacency.

On a broadcast link or NBMA link, LSAs do not need to be exchanged between DR others, and so DR others establish neighbor relationships. LSAs need to be exchanged between the DR and BDR, and between the DR, BDR, and DR others, and so they establish an adjacency with each other. In the following figure, two DR others each have three neighbor relationships and two adjacencies.



There are only OSPF adjacencies on a P2P or P2MP link.

A neighbor relationship indicates that two neighbors reach the 2-way state, whereas an adjacency indicates that two neighbors reach the Exstart state or higher.

The OSPF adjacency establishment process varies slightly depending on the network type.

On a broadcast network, the DR and BDR establish adjacencies with each router on the same network segment, but DR others establish only neighbor relationships.

On an NBMA network, all routers establish adjacencies only with the DR and BDR.

17.2.8 OSPF Area

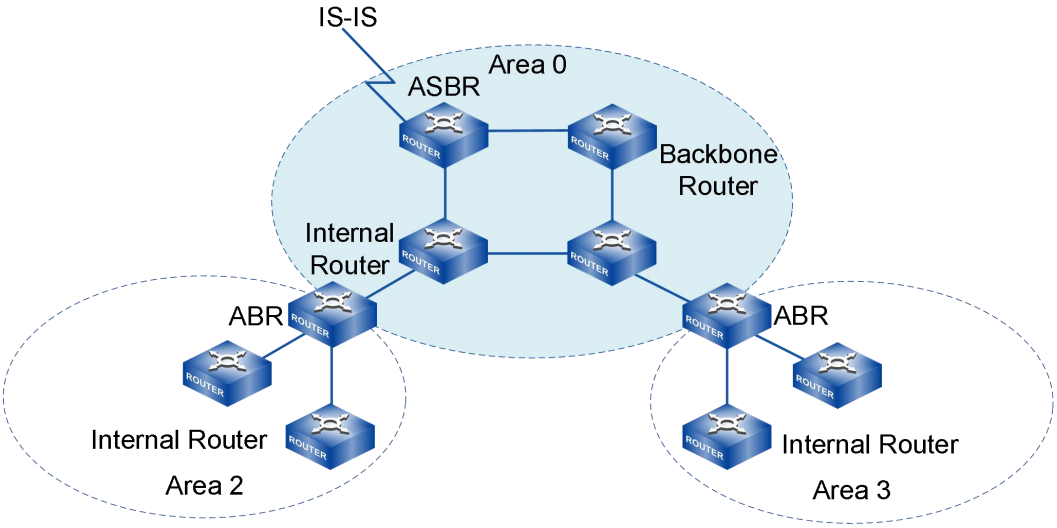
When a large number of routers run OSPF, their LSDBs become very large and require a large amount of storage space. Large LSDBs also complicate SPF calculations and increase loads on the routers. Network expansion increases the possibility of network topology changes, resulting in route flapping and frequent OSPF packet transmission. When a large number of OSPF packets are transmitted on the network, bandwidth efficiency decreases. More seriously, each change causes all the routers in the network to rerun their routing calculations.

OSPF resolves this problem by partitioning an AS into different areas. Areas are different groups of logically divided routers, each group is identified by Area ID. The boundary of the area is the router, not the link. Each OSPF interface must belong to an area, meaning that a link (or network segment) belongs to only one area.

Before learning about OSPF areas, master two concepts: router types and route types.

17.2.8.1 Router Types

The following figure shows common router types within OSPF.



Router Types	Definition
Internal Router	All interfaces on an internal router belong to the same OSPF area.
ABR (Area Border Router)	An ABR belongs to two or more areas, one of which must be a backbone area. ABR is used to connect the backbone area to the non-backbone area. It can be physically or logically connected to the backbone area.
Backbone Router	One or more interfaces on a backbone router belong to a backbone area. Internal routers in Area 0 and all ABRs are backbone routers.
Autonomous System Boundary Router (ASBR)	An ASBR exchanges routing information with other ASs. An ASBR is not required to reside on the border of an AS. It may be an internal router or an ABR. An OSPF device that has imported external routing information becomes an ASBR.

17.2.8.2 Routing Type

Inter-area and intra-area routes in an AS describe the network structure of the AS. AS external routes describe routes to destinations outside the AS. OSPF classifies imported AS external routes into either Type 1 or Type 2 external routes.

Table lists route types in descending order of priority.

Routing Type	Definition
Intra-area route	Routes within an area.
Inter-area route	Routes between areas.
Type1 External routing (Type1 External)	Type 1 external routes have high reliability. Cost of a Type 1 external route = Cost of the route from a router to an ASBR + Cost of the route from the ASBR to the destination
Type2 External routing (Type2 External)	Type 2 external routes have low reliability. Therefore OSPF considers the cost of the route from an ASBR to the destination to be much greater than the cost of any internal route to the ASBR. Cost of a Type 2 external route = Cost of the route from an ASBR to the destination

17.2.8.3 Area Types

OSPF areas include common areas, stub areas, and not-so-stubby areas (NSSAs).

Area Types	Function	Note
Common area	By default, OSPF areas are common areas. Common areas include: • Standard area: transmits intra-area, inter-area, and external routes. • Backbone area: connects to all other OSPF areas and transmits inter-area routes. The backbone area is responsible for routing between areas, and routing information between non-backbone areas must	• The backbone area must have all its devices connected. • All non-backbone areas must remain connected to the backbone area.

Area Types	Function	Note
	be forwarded through the backbone area.	
Stub area	A stub area is a non-backbone area with only one ABR and generally resides at the border of an AS. The ABR in a stub area does not transmit received AS external routes, which significantly decreases the number of entries in the routing table on the ABR and the amount of routing information to be transmitted. To ensure the reachability of AS external routes, the ABR generates a default route and advertises the route to non-ABRs in the stub area. A totally stub area allows only intra-area routes and ABR-advertised Type 3 LSAs carrying a default route to be advertised within the area.	• The backbone area cannot be configured as a stub area. • An ASBR cannot exist in a stub area. Therefore, AS external routes cannot be advertised within the stub area. • A virtual link cannot pass through a stub area.
NSSA	An NSSA is similar to a stub area. An NSSA does not advertise Type 5 LSAs. but can import AS external routes. ASBRs in an NSSA generate Type 7 LSAs to carry the information about the AS external routes. The Type 7 LSAs are advertised only within the NSSA. When the Type 7 LSAs reach an ABR in the NSSA, the ABR translates the Type 7 LSAs into Type 5 LSAs and floods them to the entire AS. A totally NSSA area allows only intra-area routes to be advertised within the area.	• ABRs in an NSSA advertise Type 7 LSAs carrying a default route within the NSSA. All inter-area routes must be advertised by ABRs. • An ABR in an NSSA advertises default routes in Type 7 LSAs within the NSSA. • All inter-area routes are advertised by ABRs. • A virtual link

Area Types	Function	Note
		cannot pass through an NSSA.

17.2.8.4 Inter-Area Loop and Loop Prevention

The shortest path first (SPF) algorithm runs within an OSPF area, which prevents routing loops within this area. The distance-vector algorithm is used for inter-area route transmission, which may easily cause routing loops.

To prevent inter-area loops, OSPF prohibits routing information from being advertised between two non-backbone areas and allows advertising routing information within a single area or between the backbone and non-backbone areas. Therefore, each ABR must be connected to the backbone area.

17.2.8.5 OSPF Default Route

Default routes have all 0s as the destination address and mask. A device uses a default route to forward packets when no matching route is discovered. Hierarchical management of OSPF routes prioritizes the default route carried in Type 3 LSAs over the default route carried in Type 5 or Type 7 LSAs.

OSPF default routes are used when:

- An ABR advertises default Type 3 LSAs instructing routers within an area to forward packets between areas.
- An ASBR advertises default Type 5 LSAs or default Type 7 LSAs instructing routers in an AS to forward packets to other ASs.

Advertising OSPF default routes must follow these rules:

- An OSPF router in an area can advertise LSAs carrying a default route only when the router has an interface connected to a device outside the area.
- If an OSPF router has advertised LSAs carrying a default route, the router no longer learns the same type of LSA advertised by other routers, which carry a default route. That is, the router uses only the LSAs advertised by itself to calculate routes. The LSAs advertised by other routers are still recorded in the LSDB.
- If a router must use a route to advertise LSAs carrying an external default route, the route cannot be a route learned by the local process. This is because external default routes guide packet forwarding outside an AS, whereas routes within an AS have the next hops pointing to the devices within the AS.

The following table lists default route advertisement principles in different areas.

Area Types	Function
Common area	By default, devices in a common OSPF area do not generate default routes. When a default route on the network is generated by another non-OSPF routing process, the device that generates the default route must advertise it within the entire OSPF AS. You can configure an ASBR to generate a default route. After configuration, the ASBR generates a Type 5 LSA carrying the default route and then advertises the LSA within the entire OSPF AS.
Stub area	A stub area does not allow AS external routes (carried in Type 5 LSAs) to be transmitted within the area. All routers within a stub area must learn AS external routes from the ABR. The ABR automatically generates a default Summary LSA (Type 3 LSA) and then advertises it within the entire stub area. This makes all routes to destinations outside an AS learnable from the ABR.
Totally stub area	A totally stub area does not allow AS external routes (carried in Type 5 LSAs) or inter-area routes (carried in Type 3 LSAs) to be transmitted within the area. All routers within the totally stub area must learn AS external routes and other areas' routes from the ABR. The ABR automatically generates a default Summary LSA (Type 3 LSA) and advertises it within the entire totally stub area. This makes all routes to destinations outside an AS and to destinations in other areas learnable from the ABR.
NSSA Area	An NSSA allows its ASBRs to import a small number of AS external routes. The ASBRs do not advertise Type 5 LSAs received from other areas within the NSSA. AS external routes can be learned only from ASBRs in the NSSA. Devices in an NSSA do not automatically generate default routes. Use either of the following methods to generate default routes: To advertise AS external routes through the ASBR in the NSSA and advertise other external routes through other areas, At the moment, a default Type 7 LSA on the ABR would be configured and advertise this LSA within the entire NSSA. In this manner, a small number of

Area Types	Function
	AS external routes can be learned from the ASBR in the NSSA, and other routes can be learned from the ABR in the NSSA. If you want all external routes to arrive only through ASBR of NSSA in this area. The ASBR must be manually configured by command, so that the ASBR can generate a default NSSA LSA(Type7 LSA) and announce it to the whole NSSA area. In this way, all external routes can only be reached through ASBR of NSSA in this area. The difference between the two methods is: An ABR generates a default Type 7 LSA regardless of whether the routing table contains the default route 0.0.0.0. An ASBR generates a default Type 7 LSA only when the routing table contains the default route 0.0.0.0. Default routes are flooded only within the local NSSA and are not flooded within the entire OSPF AS. If routers in the local NSSA cannot discover routes to the outside of the AS, the routers can forward packets outside of the AS through an ASBR. However, packets of other OSPF areas cannot be sent outside the AS through this ASBR. Default Type 7 LSAs are neither translated into default Type 5 LSAs nor flooded within the entire OSPF AS.
Totally NSSA	A totally NSSA does not allow AS external routes (carried in Type 5 LSAs) or inter-area routes (carried in Type 3 LSAs) to be transmitted within the area. All routers within the totally NSSA must learn AS external routes from the ABR. The ABR automatically generates a Type 3 LSA carrying a default route and advertises it within the entire totally NSSA. This allows all external routes received from other areas and inter-area routes to be advertised within the totally NSSA.

17.2.8.6 OSPF LSA Types

An OSPF network is partitioned into multiple areas. These areas need to maintain their own LSDBs, and routers in these areas are classified into different types. LSAs encapsulated with route descriptions can also be classified based on the router types.

- A Router-LSA is a Type 1 LSA. It describes a router's link state and cost. Every

routing device in OSPF network will issue a Type1 LSA. This type of LSA is used to describe the link state and cost of equipment, and it propagates in the area to which the router belongs.

- A Network-LSA is a Type 2 LSA. It describes the link state of all routers on the local network segment. A DR generates and advertises Network-LSAs within its area.
- A Network-summary-LSA is a Type 3 LSA. It describes inter-area routing information. An ABR generates and advertises Network-summary-LSAs within its area to notify destination addresses of routes from this area to other areas. Actually, an ABR collects Type 1 and Type 2 LSAs within its area, summarizes related routes, and then advertises these routes.
- An ASBR-summary-LSA is a Type 4 LSA. It describes routes to an ASBR. An ABR generates and advertises ASBR-summary-LSAs to the areas except the area to which the ASBR is located.
- An AS-external-LSA is a Type 5 LSA. It describes routes to a destination outside an AS. An ASBR generates and advertises AS-external-LSAs to all areas except stub areas and NSSAs.
- NSSA LSA, also called Type7 LSA. It describes routes to a destination outside the AS. An ASBR generates and advertises NSSA LSAs only in NSSAs. When an ABR in an NSSA receives NSSA LSAs, the ABR selectively translates them into Type 5 LSAs to advertise imported external routes to other areas on the OSPF network.
- Opaque LSAs include Type 9 LSAs, Type 10 LSAs, and Type 11 LSAs. They provide a universal mechanism for OSPF extensions.
  - Type 9 LSAs are advertised only on the network segment where the originating interface resides. Grace LSAs used to support graceful restart (GR) are Type 9 LSAs.
  - Type 10 LSAs are advertised inside an OSPF area. LSAs used to support traffic engineering (TE) are Type 10 LSAs.
  - Type 11 LSAs are advertised within an AS. At present, there are no applications for Type 11 LSAs.

17.2.9 OSPF Fast Convergence

OSPF fast convergence is an extended OSPF feature that speeds up route convergence. The characteristics of OSPF fast convergence are as follows:

- Priority-based OSPF convergence
  Priority-based OSPF convergence ensures that, when a great number of routes need to converge, important routes converge first. This improves network reliability. Different routes can have different convergence priorities configured.

Using priority-based OSPF convergence, you can assign higher convergence priority to routes for key services so that these routes can converge faster. This reduces performance deterioration of these key services.

- Partial Route Calculation (PRC)
  When routes on the network change, only changed routes are recalculated.
- Intelligent timer-based LSA generation and receipt. With the intelligent timer, the delay in route calculation can be flexibly configured. This allows for quick response to infrequent changes and suppresses frequent changes.
  To avoid network connections or frequent route flapping from consuming excessive device resources, RFC 2328 establishes that:
  - After an LSA is generated, it cannot be generated again within 1 second. The interval for updating LSAs is 5 seconds.
  - The interval for receiving LSAs is 1 second.

  On a stable network where routes require fast convergence, you can use the intelligent timer to set the interval for updating or receiving LSAs to 0 seconds. This ensures that topology or route changes can be immediately detected or advertised to the network, speeding up route convergence.
- Route calculation based on the intelligent timer
  When there is a change in network topology, devices recalculate routes according to OSPF. To prevent frequent network topology changes from affecting device performance, RFC 2328 requires the use of a delay timer so that route calculation is performed only after the specified delay. The delay suggested by RFC 2328 is a fixed value that is unable to ensure fast response to topology changes or effectively suppress flapping.
  The intelligent timer controls the delay in route calculation, which speeds up response to infrequent changes and effectively suppresses frequent changes.

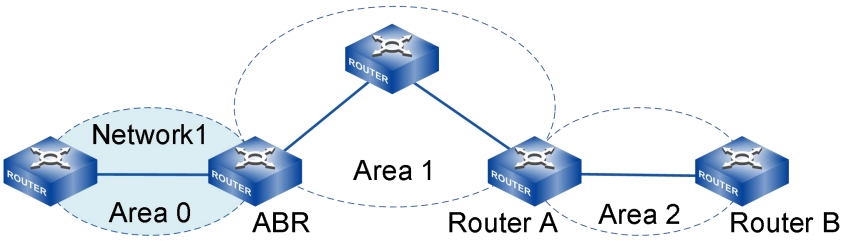
17.2.10 OSPF Virtual Link

A virtual link is a logical channel set up between two ABRs through a non-backbone area.

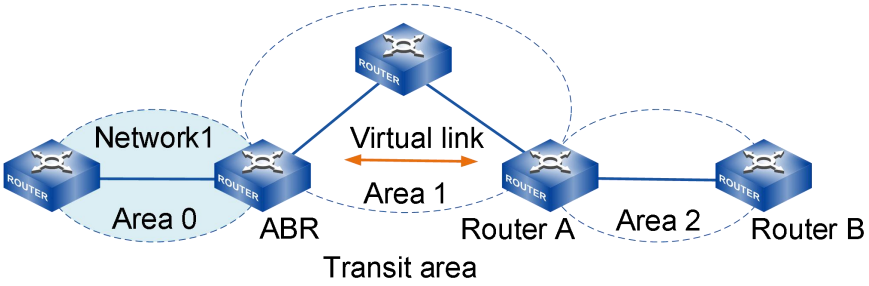
As required in RFC 2328, all non-backbone areas must be connected to the backbone area to ensure that all areas are reachable. Some non-backbone areas may not be connected to the backbone area. You can configure an OSPF virtual link to resolve this issue.

As shown in the figure below, Area 2 is not connected to the backbone area Area 0, so RouterA cannot generate routing information of Network1 in Area 0 as ABR to Area

2, so there is no route to Network1 on RouterB. A virtual link can be deployed to resolve this issue.



As shown in the figure below, OSPF message information is directly transmitted between two ABRs through virtual connection, and OSPF equipment between them only plays the role of forwarding messages. Because the destination of OSPF packets is not the OSPF device, the OSPF device transparently transmits the OSPF packets as common IP packets.



A virtual link functions as a point-to-point connection between two ABRs. The interfaces on both ends of the virtual link can be configured with parameters such as the Hello interval in the same way these parameters are configured on physical interfaces. The transmit area refers to the area that provides an internal route of a non-backbone area for both ends of the virtual link. A virtual link must be configured at both ends of the link, or it will not take effect.

However, a virtual link increases network complexity and complicates fault location. It is not recommended to use virtual links during network planning. Virtual links are only a temporary measure to fix unavoidable network topology problems. When virtual links are designed, consider whether to re-plan the network.

17.2.11 OSPF Route Summarization and Filtering

17.2.11.1 OSPF Route Summarization

Route summarization allows routes with the same prefix to be summarized into one route. Only the summarized route is advertised to other areas.

This reduces both the routing table size and routing information transmitted between areas, thereby improving device performance.

Route summarization can be carried out by either an ABR or ASBR following these conditions:

- ABR

  An ABR generates Network-summary (Type 3) LSAs by network segment while advertising routing information to other areas. If the network segments are contiguous, the ABR can summarize them into a single segment. This allows the ABR to send one summarized LSA for the specified network segments.
- ASBR

  An ASBR can summarize imported AS-external (Type 5) LSAs within a summarized address range. After an NSSA is configured, the ASBR must summarize imported NSSA (Type 7) LSAs with the summarized address range. Devices that are both ABR and ASBR summarize Type 5 LSAs translated from Type 7 LSAs.

17.2.11.2 OSPF Route Filtering

OSPF supports route filtering using routing policies. By default, OSPF does not filter routes.

These policies include route-policy, access-list, and prefix-list.

OSPF route filtering can be used to:

- Import routes.

  OSPF can import routes learned by other routing protocols. You can configure routing policies to filter imported routes, allowing OSPF to import only routes that match specific conditions.
- Advertise imported routes.

  OSPF advertises imported routes to its neighbors.

  You can configure filtering rules to filter routes to be advertised. Filtering rules can be configured only on ASBRs.
- Learn routes.

  Filtering rules can be configured to allow OSPF to filter received intra-area, inter-area, and AS external routes.

  After receiving routes, an OSPF device adds only the routes that match filtering rules to the local routing table. However, the device can advertise all routes from the OSPF routing table.
- Learn inter-area LSAs.

  You can configure an ABR to filter incoming Summary LSAs. This configuration

takes effect only on ABRs because only ABRs can advertise Summary LSAs.

The differences between inter-area LSA learning and route learning: Inter-area LSA learning filters incoming LSAs. However, Route learning filters the routes that are calculated based on LSAs, but does not filter LSAs. All incoming LSAs can be learned in route learning.

- Advertise inter-area LSAs.

  You can configure an ABR to filter outgoing Summary LSAs. This configuration takes effect only on ABRs.

17.2.12 OSPF Multi-Process

OSPF supports multi-processes, and many different OSPF processes can run on the same router, which do not affect each other and are independent of each other. Routing interactions between different OSPF processes are equivalent to routing interactions between different routing protocols.

An interface of a router can only belong to one OSPF process.

A typical application of OSPF multiprocess is to run OSPF protocol between PE and CE in VPN scene, and IGP on VPN backbone also adopts OSPF. On PE, these two OSPF processes do not affect each other.

17.2.13 OSPF RFC 1583 Compatibility

RFC 1583 is an early version of OSPFv2.

During OSPF external route calculations, routing loops may occur because RFC 2328 and RFC 1583 define different route selection rules. To prevent routing loops, both communication ends must use the same route selection rules.

- When RFC 1583 compatibility is enabled, OSPF uses route selection rules defined in RFC 1583.
- When RFC 1583 compatibility is disabled, OSPF uses route selection rules defined in RFC 2328.

OSPF calculates external routes based on Type 5 LSAs. If the router with RFC 1583 compatibility receives a Type 5 LSA:

- The router selects a route to the ASBR that originates the LSA or to the forwarding address (FA) described in the LSA.
- The router selects external routes to the same destination.

OSPF is not compatible with RFC1583 by default.

17.3 OSPF Configuration

17.3.1 Enter OSPF Router Configuration View

【Command】

router ospf [<PROCESS-ID>]

no router ospf [<PROCESS-ID>]

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

PROCESS-ID: process ID, range 1-65535, default process 0.

【Description】

router ospf [PROCESS-ID]: the command is used to start the OSPF process and enter OSPF process mode.

【Instance】

```
# configuration starts and enters the OSPF process 0
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
```

17.3.2 OSPF Route ID

【Command】

router-id <*A.B.C.D*>

no router-id [<*A.B.C.D*>]

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D：router ID.

【Description】

router-id [< A.B.C.D>] : the command is used to configure the OSPF router ID of the device.

The same configuration command router-id <A.B.C.D> is available in configuration mode with a lower priority

【Instance】

```
# Configure device router ID to 1.1.1.1
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#router-id 1.1.1.1
```

17.3.3 OSPF Area and Enablement

【Command】

network (A.B.C.D/M | A.B.C.D) area (A.B.C.D | <0-4294967295>)
no network (A.B.C.D/M | A.B.C.D) area (A.B.C.D|<0-4294967295>)

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D/M | A.B.C.D: network address and mask.

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

【Description】

network: the command is used to configure the OSPF region the various network interfaces of the device belong to and enable OSPF functions.

By default, the interface does not belong to any region and OSPF function is disabled.

【Instance】

```
# specifies that the primary IP address of the interface running the OSPF protocol is on the network segment 10.11.20.0/24, and the OSPF region ID of the interface is 1
Switch> enable
Switch#configure terminal
```

```
Switch(config)#router ospf
Switch(config-router)#network 10.11.20.0 0.0.0.255 area 1
Or Switch(config- router)#network 10.11.20.0/24 area 1
```

17.3.4 Configure and Publish a Host Route

【Command】

```
host (A.B.C.D) area (A.B.C.D | <0-4294967295>) [cost <COST>]
no host (A.B.C.D) area (A.B.C.D | <0-4294967295>) [cost <COST>]
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: host address.

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

<cost> : host routing cost value, integer in range 0~65535, default value is 0.

【Description】

host: the command is used to configure and publish a host route.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#host 172.16.10.100 area 1
```

17.3.5 Configure Neighbor Interface Address in the NBMA Network

【Command】

```
neighbor <A.B.C.D> [priority <PRIORITY-ID>] [poll-interval
<SECONDS>] [cost <METRIC>]
no neighbor <A.B.C.D> [priority <PRIORITY-ID>] [poll-interval
<SECONDS>] [cost <METRIC>]
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: neighbor interface address.

PRIORITY-ID: priority of neighbors in ID network, invalid in P2MP network, default value is 0, value range is <0-255>.

SECONDS: interval of HELLO package before establishing neighborhood relationship, value range is <1-65535>.

METRIC: neighborhood link cost, value range is <1-65535>.

【Description】

neighbor: the command is used to configure neighbors in the NBMA network.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#neighbor 10.11.20.1
```

17.3.6 Create Virtual Connection

【Command】

```
area (A.B.C.D|<0-4294967295>) virtual-link A.B.C.D
[authentication] [authentication-key LINK] [dead-interval
<1-65535>] [hello-interval <1-65535>] [message-digest-key <1-255>]
[retransmit-interval <1-65535>] [transmit-delay <1-65535>]
no area (A.B.C.D|<0-4294967295>) virtual-link A.B.C.D
[authentication] [authentication-key LINK] [dead-interval
<1-65535>] [hello-interval <1-65535>] [message-digest-key <1-255>]
[retransmit-interval <1-65535>] [transmit-delay <1-65535>]
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

A.B.C.D: virtual link neighbor router ID, IP address format.

hello-interval seconds: the time interval for the interface to send hello message, ranging from 1 to 65535 in seconds. The default value is 10 seconds. This value must equal the value of hello seconds on its connected virtual connection router.

retransmit-interval seconds: the time interval for retransmission of LSA packets of the interface, ranging from 1 to 65535 in seconds. The default value is 5 seconds.

transmit-delay seconds: the time interval for transmitting LSA packets, ranging from 1 to 65535 in seconds. The default value is 1 second.

dead-interval seconds: dead interval seconds, ranging from 1 to 65535 in seconds, with a default value of 40 seconds. This value must equal the dead-interval seconds value of its virtual connected router and be at least 4 times the hello seconds value.

authentication: plaintext authentication mode.

message-digest: MD5 authentication mode, it is optional, if unselected it will be simple authentication mode (password in plaintext).

null: null is unauthenticated.

authenticate-key: plaintext authentication password.

message-digest-key: MD5 key range 1-255.

【Description】

virtual-link: this command is used to create and configure a virtual connection.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#area 1 virtual-link 1.1.1.1
```

17.3.7 Routing Within the Aggregation Area

【Command】

```
area (A.B.C.D|<0-4294967295>) range (A.B.C.D/M) [advertise | not-advertise]
no area (A.B.C.D|<0-4294967295>) range (A.B.C.D/M) [advertise | not-advertise]
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

A.B.C.D/M: aggregate address and mask.

advertise | not-advertise: whether to publish the aggregation route or not. This aggregation route is published if the parameter advertise or not-advertise is not specified.

【Description】

area range: the command is used to aggregate the routers of an area. If the network addresses in the region are configured in a discontinuous manner, configuring this command on ABR notifies a summary route that contains all the individual networks in the region that belong to a particular range.

By default this function is not configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#area 1 range 10.11.20.0/24
```

17.3.8 Type-3 LSA Filter

【Command】

```
area (A.B.C.D|<0-4294967295>) filter-list (access|prefix) <NAME>
(in|out)
no area area (A.B.C.D|<0-4294967295>) filter-list (access|prefix)
<NAME> (in|out)
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

access：specifies to use access-list.

Prefix: specifies to use prefix-list.

<NAME>: specified Access-list/prefix-list name/ID, length 1-256.

in: from other areas to this area.

out: from this area to other areas.

【Description】

area filter-list: the command is used to configure the filtering of Type-3 LSA in and out of the area.

By default this function is not configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 1 deny 172.22.0.0 0.0.0.255
Switch(config)#access-list 1 permit any
Switch(config)#router ospf
Switch(config- router)#area 1 filter-list access 1 in
```

17.3.9 Stub Area

【Command】

```
area (A.B.C.D|<0-4294967295>) stub [no-summary]
no area (A.B.C.D|<0-4294967295>) stub [no-summary]
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

no-summary: this parameter is only used for the ABR of the Stub region. After configuration, the ABR only publishes a type-3 LSA for the default route to the Stub region, and does not generate any other type-3 LSAs (this region is also known as the Totally Stub region).

【Description】

area stub: the command is used to configure a region as a Stub region.

By default, no region is set to Stub region.

If user want to configure a zone as a Stub zone, all routers in the zone must configure this property. For configuration, refer to the command default-cost.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#area 1 stub
```

17.3.10 NSSA Area

【Command】

```
area (A.B.C.D|<0-4294967295>) nssa
[default-information-originate [always | metric <0-16777214> |
metric-type <1-2>] | translator-role (always | candidate | never)
| no-summary | no-redistribution]
no area (A.B.C.D|<0-4294967295>) nssa
[default-information-originate | translator-role | no-summary |
no-redistribution]
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

default-information-spanish [metric <0-16777214>] [metric <1-2>] : generates LSA of type-7 metric <0-16777214> specifies the metric value, and metric-type <1-2> specifies the metric type of external-LSA, 2 is the default value.

translator-role {candidate|never|always} : specify LSA transformation behavior of router:

- candidate: means that if the router is elected as translator, NSSA-LSA can be converted to Type-5 LSA, which is candidate by default.
- never: means that the router never converts NSSA-LSA to Type 5 LSA.

- always: means that the router always converts NSSA-LSA to Type 5 LSA.

no-summary: This parameter is used only for the ABR of the NSSA region. After configuration, NSSA ABR publishes a default route to the region only through the Summary-LSA of Type-3, and does not publish any other summary-LSAS to the region (this region is also known as the Totally NSSA region).

no-redistribution: means do not distribute EXTERNAL - LSA to NSSA.

【Description】

area nssa: the command is used to configure a region as a nssa region.

By default, no region is set to nssa region.

If user want to configure a zone as a NSSA zone, all routers in the zone must configure this command. For configuration, refer to the command default-cost.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#area 1 nssa
```

17.3.11 Area Shortcut Mode

【Command】

area (A.B.C.D|<0-4294967295>) **shortcut {default|enable|disable}**

no area (A.B.C.D|<0-4294967295>) **shortcut {default|enable|disable}**

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

default: set the default shortcut behavior of the area.

enable: forced to shortcut when through the area.

disable: not to screenshot when through the area.

Default: the default configuration is set to default.

【Description】

area shortcut: the command is used to configure shortcut mode of a region.

Whether the region boundary router is connected to a backbone router or not, enabling region shortcut allows traffic to pass through non-backbone regions with lower magnitude.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#area 1 shortcut enable
```

17.3.12 Default Costs for Introduced Default Routes

【Command】

```
area (A.B.C.D|<0-4294967295>) default-cost <cost>
no area (A.B.C.D|<0-4294967295>) default-cost
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

<COST>:value range is 0-16777215. Default value is 1.

【Description】

area default-cost: this command used to configure stub/nssa zone to introduce the default cost of default routing.

This command only applies to an ABR router connected to a stub region or NSSA region.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#area 1 default-cost 20
```

17.3.13 Configure Aggregation of External Routes and Notify or Suppress

【Command】

summary-address <A.B.C.D/M> **[{not-advertise | tag** <TAG-VALUE>**}]**

no summary-address <A.B.C.D/M> **[{not-advertise | tag** <TAG-VALUE>**}]**

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D/M: aggregate address and mask.

not-advertised: control the external routing.

tag <TAG-VALUE> : the tag number of the external route, the value range is 0~4294967295, the default is 0.

【Description】

summary-address: this command is used to introduce OSPF from other routing protocols, demanding that each route be notified separately in an external LSA. Using this command, only one summary route can be notified for introduced routes that are covered by specific network addresses and masks, which can greatly reduce the size of the link-state database.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#summary-address 10.11.0.0/16
```

17.3.14 Link Cost Bandwidth Reference Value

【Command】

auto-cost reference-bandwidth <BANDWIDTH>

no auto-cost reference-bandwidth

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

<BANDWIDTH > : the reference bandwidth value in megabits per second, ranging from 1~4294967.

【Description】

reference-bandwidth: this command is used to configure the bandwidth reference values by which link cost is calculated.

By default, link cost is calculated based on a bandwidth reference value of 100Mbps.

If the link cost value configuration is not displayed, OSPF calculates the cost based on link bandwidth (cost = bandwidth reference value ÷ bandwidth, the maximum cost is 65535 when the calculated cost value is greater than 65535).

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#auto-cost reference-bandwidth 50
```

17.3.15 Opaque LSA Publishing and Receiving Capabilities

【Command】

capability opaque

no capability opaque

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

capability opaque: the command is used to enable OSPF opaque LSA to publish and receive Type9, Type10, and Type11 opaque LSA.

The no capability opaque: this command is used to restore the default value.

By default, opaque LSA publish receiving capability of OSPF is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#capability opaque
```

17.3.16 Compatibility with RFC1583

【Command】

```
compatible rfc1583
no compatible rfc1583
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

compatible rfc1583: the command is used to enable the compatibility of RFC 1583 with external routing rules.

no compatible rfc1583: the command is used to disable the compatibility of RFC 1583 with external routing rules.

By default, not enable the compatibility of RFC 1583 with external routing rules.

When multiple AS-External-LSA publish routes to the same destination address, the priority rules defined by RFC 1583 and RFC 2328 differ in how to select the optimal route. When RFC 2328 is compatible with RFC 1583, the regional routing in the backbone area is optimized. When RFC 2328 is incompatible with RFC 1583, intra-region routing in non-backbone area is preferred to minimize the burden of backbone area.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#compatible rfc1583
```

17.3.17 Introduce Default Route

【Command】

default-information originate [always | metric | metric-type | route-map]

no default-information originate [always | metric | metric-type | route-map]

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

always: notify default routing whether or not there is one in the software.

metric = metric <value> : sets the metric used to create the default route, and <value> ranges from 0 to 16777214. The default metric is 1.

Metric -type = metric-type {1|2} :sets the OSPF external link type of the default route.

- 1: Set the external type 1 metric of OSPF.
- 2: Set the external type 2 metric of OSPF.

route-map = route-map <WORD> : <WORD> specifies the policy name of the applied routing. Only if the default route exists in the routing table of the current router and the route matches the routing policy specified by route-map can a Type-5 LSA describing the default route be published, and the specified routing policy will affect the value in the Type-5 LSA. If the always parameter is specified at the same time, as long as a route matches the specified routing policy, a Type-5 LSA describing the default route will be published regardless of whether there is a default route in the routing table of the current router, and the specified routing policy will affect the value in the Type-5 LSA.

【Description】

default-information originate: the command is used to introduce the default route into the OSPF routing area.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#default-information originate always
```

17.3.18 OSPF Routing Default Metric

【Command】

default-metric < METRIC >
no default-metric

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

<METRIC> : sets the default metric when routing is introduced, value ranging from 0 to 16777214.

【Description】

default-metric: sets the default metric for the OSPF routing protocol, and the no command restore to the default values. When metric are incompatible due to default, routing introductions still can continue. If the metric cannot be converted, the default metric provides an alternative option that enables the routing introduction to proceed. This command causes the current routing protocols to use the same metric for all introduced routes. This command should be used in combination with the command redistribute.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#default-metric 100
```

17.3.19 OSPF Route Management Distance

【Command】

distance { <VALUE> | **ospf** }
no default { <VALUE> | **ospf** }

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

< value > : OSPF route management distance value, value range 1~255.

ospf { external | inter-area | intra-area }:

- external <external distance>: sets the management distance value of the routes learned from other routing fields, <external distance> management distance value, and its value range is 1~255.
- inter-area <inter-distance>: sets the value of route management distance from one area to another. <inter-distance> management distance value, which ranges from 1 to 255.
- intra-area <intra-distance>: sets the managed distance value of all routes in an area <intra-distance> management distance value, which ranges from 1 to 255.

【Description】

distance: the command is used to set the OSPF route management distance based on the route type configuration, and the no command restores the default value.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#distance ospf inter-area 20 intra-area 10
external 40
```

17.3.20 Ingress Route Information Filter

【Command】

```
distribute-list <ACCESS-LIST-NAME> in
no distribute-list <ACCESS-LIST-NAME> in
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

<ACCESS-LIST-NAME> : specifies the name of the access-list used.

【Description】

distribute-list in: the command is used to filter routing information in the incoming direction, and can not filter the LSA.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config-router)#distribute-list 10 in
```

17.3.21 Egress Route Information Filter

【Command】

distribute-list <ACCESS-LIST-NAME> out { connected | static | rip | bgp}

no distribute-list <ACCESS-LIST-NAME> out { connected| static | rip | bgp}

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

<ACCESS-LIST-NAME> : specifies the name of the access-list used.

out: routing updates sent by filter.

connected：connected route

static: static route;

rip：RIP route.

bgp：BGP route.

【Description】

distribution-list out: The command is used to filter outgoing routes when routes from other routing protocols are introduced into the OSPF routing table, and this command is used only by ASBR.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 11 permit 172.10.0.0 0.0.255.255
Switch(config)#router ospf
Switch(config-router)#distribute-list 1 out bgp
Switch(config-router)#redistribute bgp
```

17.3.22 Introduce Additional Routing Information

【Command】

redistribute {connected|static|rip|bgp} [metric <VALUE>**]** **[metric-type {1|2}] [route-map** <WORD>**] [tag** <TAG-VALUE>**]**

no redistribute {connected|static|rip|bgp} [metric <VALUE>**]** **[metric-type {1|2}] [route-map** <WORD>**] [tag** <TAG-VALUE>**]**

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

connected：connected route

static: static route;

rip：RIP route.

bgp：BGP route.

metric: <VALUE> the metric assigned to the introduced route, with a value range of 0-16777214.

metric -type: {1|2} the introduced metric type of the external route and can only be 1 or 2, default value is 2.

route-map: < WORD> points to the route map used to introduce the route.

tag: <tag-value> the tag number of the external route, the value range is 0~4294967295, the default value is 0.

【Description】

redistribute: the command is used only by ASBR to introduce routes from other routing protocols into the OSPF routing table.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#redistribute bgp metric 12
```

17.3.23 Interface Hello Message Control

【Command】

passive-interface <IFNAME> [<IP-ADDRESS>]

no passive-interface <IFNAME> [<IP-ADDRESS>]

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

<IFNAME>: interface name.

<IP-ADDRESS>: interface IP address.

【Description】

passion-interface: the command is used to configure not to send hello messages on a specific interface.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config-router)#passive-interface vlanif2
```

17.3.24 SPF Timer

【Command】

timers spf <SPF-DELAY> <SPF-HOLDTIME>

no timers spf

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

<SPF-DELAY>: the default is 5 seconds, and the value range is 0-2147483647.

<SPF-HOLDTIME>: the default is 10 seconds, and the value range is 0-2147483647.

【Description】

timbers spf: the command is used to configure the delay time between receiving a topological change and SPF computation, and the hold time between two discrete SPF computations.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#timers spf 6 12
```

17.3.25 The Maximum Number of LSA

【Command】

overflow database <MAXDBSIZE > **[{hard|soft}]**

no overflow database

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

< MAXDBSIZE > : maximum number of LSA, ranging from 0 to 4294967294.

soft: soft limit, a warning will be sent when crossing a line.

hard: hard limit, the ospf instance will be close directly when crossing the line. By default, it is processed as hard.

【Description】

overflow database: the command is used to configure the maximum number of LSA. no operation does not limit the maximum number.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#overflow database 10000
```

17.3.26 Maximum Concurrent Number of Current DD Packet

【Command】

```
max-concurrent-dd <VALUE>
no max-concurrent-dd
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

<VALUE> : the value range is <1-65535>, the upper limit of the process of maximum concurrent DD packet.

【Description】

max-concurrent-dd: the command is used to configure the OSPF course to process the maximum number of concurrent DD currently, and the no command restores the default value.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#max-concurrent-dd 25
```

17.3.27 Regional OSPF Message Validation

【Command】

```
area (A.B.C.D|<0-4294967295>) authentication [message-digest]
no area (A.B.C.D|<0-4294967295>) authentication
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: area identification, IP address format.

<0-4294967295> : area identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

message-digest: MD5 authentication mode, it is optional, if unselected it will be simple authentication mode (password in plaintext).

【Description】

area-id authentication: the command is used to configure the OSPF zone to authenticate OSPF messages and the authentication mode.

no area-id authentication: the command is used to cancel the configured authentication mode for this area.

By default, regions do not authenticate OSPF packets

【Instance】

```
#Configure OSPF zone 0 to use MD5 authentication mode
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config- router)#area 0 authentication message-digest
Switch(config- router)#no area 0 authentication
```

17.3.28 OSPF ABR Type

【Command】

```
ospf abr-type (cisco | ibm | shortcut | standard)
no ospf abr-type
```

【View】

OSPF View

【Default Level】

2: Configuration level

【Parameter】

Cisco: using the cisco (RFC3509) ABR implementation, the router connects multiple active areas, one of which is the backbone area.

ibm: using the ibm (RFC3509) ABR implementation, the router connects multiple active areas and backbone areas, backbone areas should not be the active areas.

shortcut: using standard (draft-ietf-ospf-shortcut-abr-02) specify a shortcut-ABR.

standard: using the standard (RFC2328) ABR implementation, the router connects multiple active areas.

【Description】

ospf abr-type: The command is used to configure the ABR type of the OSPF region.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router ospf
Switch(config-router)# ospf abr-type standard
```

17.3.29 Interface OSPF Message Validation

【Command】

ip ospf authentication [message-digest | null]
no ip ospf authentication

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

message-digest: MD5 authentication mode, optional, unselected simple authentication mode (password in plaintext).

null: null is unauthenticated.

【Description】

ip ospf authentication [message-digest]: the command is used to configure the interface to authenticate OSPF messages and the authentication mode.

ip ospf authentication null and no ip ospf authentication: the commands are used to cancel the authentication mode configured by the associated interface.

By default, relative interface do not authenticate OSPF packets

【Instance】

```
#Configure vlanif1 to use plaintext authentication mode
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip ospf authentication
Switch(config- vlanif1)#no ip ospf authentication
```

Notice

- The authentication mode of the interface has a higher priority than the authentication mode of the region. When the authentication mode is not configured in the interface and the regional configuration is simple plaintext password authentication, the password needs to be set using ip ospf authentication-key command in the interface mode. If it is MD5 authentication, user need to configure the MD5 key using ip ospf message-digest-key in interface mode. When the authentication mode of an interface is configured, the authentication of the region in which the interface is located cannot affect the authentication behavior of the interface.
- Priority: ip ospf authentication is higher than area <area-id> authentication.
- Ip ospf authentication [message-digest] and other commands of the layer 2 interface are invalid.

17.3.30 Interface OSPF Message Authentication Key

【Command】

```
ip ospf authentication-key <LINE>
no ip ospf authentication-key
ip ospf message-digest-key <KEY-ID> md5 <LINE>
no ip ospf message-digest-key <KEY-ID>
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

LINE: plaintext authentication password, no more than 8 characters; MD5 authentication password, no more than 16 characters.

KEY-ID: MD5 key ID range 1-255.

【Description】

ip ospf authentication-key <LINE >: the command is used to configure the password of the relevant interface for plaintext encryption authentication of OSPF messages, and restore the default value through the no operation of this command.

ip ospf message-digest-key <KEY-ID> md5 <LINE>: the command is used to configure the key value and password of the relevant interface for MD5 encryption verification of OSPF messages, and restore the default value through the no operation of this command.

By default, the relative interfaces do not configure passwords and keys.

【Instance】

```
# Configure vlanif1 to use MD5 authenticated key values and passwords.
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip ospf message-digest-key 2 md5 yourpassword
Switch(config- vlanif1)#no ip ospf message-digest-key 2
```

Notice

Use MD5 encryption authentication to achieve security between OSPF routers on the network. This command must configure the same key id and key between neighbors, otherwise an adjacency relationship cannot be established. The subsequent configuration of this command covers the previous configuration to prevent the system from continuing to communicate using the previous key id.

17.3.31 Interface OSPF Cost

【Command】

ip ospf cost <COST>
no ip ospf cost

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

<COST>:cost, value range is 1-65535.

【Description】

ip ospf cost: the command is used to configure the cost required to run the ospf protocol on the interface. The no operation of this command is to restore the default value.

【Instance】

```
Switch> enable
```

```
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip ospf cost 20
```

17.3.32 Interface LSA Database Filter

【Command】

```
ip ospf database-filter all out
no ip ospf database-filter
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

Ip ospf database-filter all out: the command is used to configure to enable the LSA database filter switch on a specific interface.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip ospf database-filter all out
```

17.3.33 Interface OSPF Neighbor Failure Time

【Command】

```
ip ospf dead-interval <TIME>
no ip ospf dead-interval
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

<TIME>: OSPF neighbor failure time, in seconds, value range 1 ~ 65535

【Description】

Ip ospf dead-interval: the command is used to configure the OSPF neighbor failure time.

By default, the failure time of OSPF neighbors of P2P and Broadcast interfaces is 40 seconds. The failure time of OSPF neighbors of P2MP and NBMA interfaces is 120 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip ospf dead-interval 60
```

17.3.34 Time Interval of Sending Hello Message

【Command】

ip ospf hello-interval <TIME>

no ip ospf hello-interval

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

<TIME> : the time interval for the interface to send Hello message, in seconds, with a value range of 1 ~ 65535.

【Description】

ip ospf hello-interval: the command is used to configure the time interval for the interface to send the Hello message.

By default, the time interval for P2P and Broadcast type interfaces to send Hello message is 10 seconds. The time interval for P2MP and NBMA type interface to send Hello message is 30 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip ospf hello-interval 20
```

17.3.35 Interface Retransmission LSA Interval

【Command】

```
ip ospf retransmit-interval <TIME>
no ip ospf retransmit -interval
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

<TIME> : the time interval for the interface to retransmit LSA, in seconds, with a value range of 1 ~ 65535.

【Description】

ip ospf retransmit-interval command is used to configure the time interval for the interface to retransmit LSA.

By default, the time interval for interface to retransmit LSA is 5 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip ospf retransmit-interval 10
```

17.3.36 Interface to LSA Transmission Delay Time

【Command】

```
ip ospf transmit-delay <TIME>
no ip ospf transmit-delay
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

<TIME> : transmission delay time of LSA of the interface, in seconds, with a value range of 1 ~ 65535.

【Description】

ip ospf retransmit-delay: the command is used to configure the transmission delay of LSA of the interface.

By default, the transmission delay for interface to retransmit LSA is 1 second.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip ospf transmit-delay 2
```

17.3.37 Disable the OSPF Function of the Interface

【Command】

ip ospf disable all

no ip ospf disable all

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip ospf disable all: disable OSPF function on the interface.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip ospf disable all
```

17.3.38 Interface OSPF MTU Value

【Command】

ip ospf mtu <MTU>

no ip ospf mtu

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

< MTU>: interface MTU value, ranging from 576 to 65535.

【Description】

ip ospf mtu: the command is used to configure the MTU value of the interface as the basis for the OSPF grouping. The interface values configured by this command are used only by the ospf protocol and are not updated to the kernel.

By default, the interface MTU value obtained from the kernel is used.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip ospf mtu 1480
```

17.3.39 Interface DD Exchanges Ignore MTU

【Command】

```
ip ospf mtu-ignore
no ip ospf  mtu-ignore
```

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

ip ospf mtu-ignore: the command does not check the MTU size when configuring the DD switching for the interface. By default, check is necessary.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
```

```
Switch(config-vlanif1)#ip ospf mtu-ignore
```

17.3.40 Interface OSPF Network Type

【Command】

ip ospf network {broadcast | non-broadcast | point-to-point | point-to-multipoint}

no ip ospf network

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

broadcast: set OSPF network type to broadcast.

non-broadcast: set OSPF network type to NBMA.

point-to-point: set OSPF network type to point-to-point.

point-to-multipoint: set OSPF network type to point-to-multipoint.

【Description】

ip ospf network: the command is used for the OSPF network type of the interface.

By default, the OSPF network type of the interface is the broadcast type.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip ospf network non-broadcast
```

17.3.41 The DR Priority of the Interface

【Command】

ip ospf priority <PRIORITY>

no ip ospf priority

【View】

Layer 3 Interface View

【Default Level】

2: Configuration level

【Parameter】

<PRIORITY> : DR priority of the interface, ranging from 0 to 255.

【Description】

ip ospf priority: the command is used to configure the DR priority of the interface.

By default, the DR priority of the interface is 1.

The DR priority determines the qualification of the interface for election of DR/BDR. The higher the value, the higher the priority. High priority will be taken into account when voting rights conflict. If a device has a priority of 0, it will not be elected as DR or BDR.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config- vlanif1)#ip ospf priority 2
```

17.3.42 Display OSPF Route Information

【Command】

show ip ospf route

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show ip ospf route: the command is used to display the routing information learned by ospf protocol.

The show ip route ospf command is used to display the route information shown in the routing table that learned by ospf protocol.

【Instance】

```
SwitchA> enable
SwitchA#configure terminal
SwitchA(config)#interface vlanif1
```

```
SwitchA(config- vlanif1)#ip address 192.168.1.1/24
SwitchA(config- vlanif1)#exit
SwitchA(config)#vlan database
SwitchA(config-vlan)#vlan 2
SwitchA(config-vlan)#exit
SwitchA(config)#ip interface vlan 2
SwitchA(config)#interface vlanif2
SwitchA(config- vlanif2)#ip address 192.168.2.2/24
SwitchA(config- vlanif2)#exit
SwitchA(config)#router ospf
SwitchA(config-router)#network 192.168.1.0/24 area 0
SwitchA(config-router)#network 192.168.2.0/24 area 0

SwitchB> enable
SwitchB#configure terminal
SwitchB(config)#vlan database
SwitchB(config-vlan)#vlan 2
SwitchB(config-vlan)#exit
SwitchB(config)#ip interface vlan 2
SwitchB(config)#interface vlanif2
SwitchB(config- vlanif2)#ip address 192.168.2.1/24
SwitchB(config- vlanif2)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#network 192.168.2.0/24 area 0
SwitchB(config-router)#end
Switch#show ip ospf route
O  192.168.1.0/24 [2] via 192.168.2.2, vlanif2, Area 0.0.0.0
C  192.168.2.0/24 [1] is directly connected, vlanif2, Area 0.0.0.0

Switch#show ip route  ospf
O      192.168.1.0/24 [110/2] via 192.168.2.2, vlanif2, 00:26:23
```

17.3.43 Display OSPF Database Information

【Command】

show ip ospf database

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show ip ospf database: the command is used to display the routing status database information of the ospf protocol.

【Instance】

```
SwitchA> enable
SwitchA#configure terminal
SwitchA(config)#interface vlanif1
SwitchA(config- vlanif1)#ip address 192.168.1.1/24
SwitchA(config- vlanif1)#exit
SwitchA(config)#vlan database
SwitchA(config-vlan)#vlan 2
SwitchA(config-vlan)#exit
SwitchA(config)#ip interface vlan 2
SwitchA(config)#interface vlanif2
SwitchA(config- vlanif2)#ip address 192.168.2.2/24
SwitchA(config- vlanif2)#exit
SwitchA(config)#router ospf
SwitchA(config-router)#network 192.168.1.0/24 area 0
SwitchA(config-router)#network 192.168.2.0/24 area 0

SwitchB> enable
SwitchB#configure terminal
SwitchB(config)#vlan database
SwitchB(config-vlan)#vlan 2
SwitchB(config-vlan)#exit
SwitchB(config)#ip interface vlan 2
SwitchB(config)#interface vlanif2
SwitchB(config- vlanif2)#ip address 192.168.2.1/24
SwitchB(config- vlanif2)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#network 192.168.2.0/24 area 0
SwitchB(config-router)#end
SwitchB#show ip ospf database

                Router Link States (Area 0.0.0.0)

Link ID         ADV Router      Age  Seq#       CkSum  Link count
192.168.2.1     192.168.2.1     570  0x80000003 0xe0bd 1
```

```
192.168.2.2     192.168.2.2      552 0x80000005 0xf825 2

                Net Link States (Area 0.0.0.0)

Link ID         ADV Router      Age  Seq#       CkSum
192.168.2.2     192.168.2.2      571 0x80000001 0x881d
```

17.3.44 Display OSPF Neighbor Information

【Command】

show ip ospf neighbor

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

show ip ospf neighbor: the command is used to display neighbor information of the ospf protocol.

【Instance】

```
SwitchA> enable
SwitchA#configure terminal
SwitchA(config)#interface vlanif1
SwitchA(config- vlanif1)#ip address 192.168.1.1/24
SwitchA(config-vlanif1)#exit
SwitchA(config)#vlan database
SwitchA(config-vlan)#vlan 2
SwitchA(config-vlan)#exit
SwitchA(config)#ip interface vlan 2
SwitchA(config)#interface vlanif2
SwitchA(config- vlanif2)#ip address 192.168.2.2/24
SwitchA(config)#router ospf
SwitchA(config-router)#network 192.168.1.0/24 area 0
SwitchA(config-router)#network 192.168.2.0/24 area 0

SwitchB> enable
SwitchB#configure terminal
```

```
SwitchB(config)#vlan database
SwitchB(config-vlan)#vlan 2
SwitchB(config-vlan)#exit
SwitchB(config)#ip interface vlan 2
SwitchB(config)#interface vlanif2
SwitchB(config- vlanif2)#ip address 192.168.2.1/24
SwitchB(config-vlanif2)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#network 192.168.2.0/24 area 0
SwitchB(config-router)#end
SwitchB#show ip ospf neighbor

OSPF process 0:
Neighbor ID      Pri    State              Dead Time    Address
Interface
192.168.2.2      1    Full/DR          00:00:37    192.168.2.2
vlanif2
```

18 BGP Configuration

18.1 Overview

The Border Gateway Protocol (BGP) is a path vector protocol that allows devices between Autonomous Systems (ASs) to communicate and selects optimal routes. BGP-1 (defined in RFC 1105), BGP-2 (defined in RFC 1163), and BGP-3 (defined in RFC 1267) are three earlier versions of BGP. BGP-4 (defined in RFC 1771) has been used since 1994. Since 2006, unicast IPv4 networks have been using BGP-4 defined in RFC 4271, and other networks (such as IPv6 networks) have been using Multiprotocol BGP (MP-BGP) defined in RFC 4760.

MP-BGP is an extension of BGP-4 and applies to different networks; however, the original message exchange and routing mechanisms of BGP-4 are not changed. MP-BGP applications on IPv6 unicast and IPv4 multicast networks are called BGP4+ and Multicast BGP (MBGP) respectively.

A network is divided into different ASs to facilitate network management. In 1982, the Exterior Gateway Protocol (EGP) was used to dynamically exchange routing information between ASs. However, EGP advertises only reachable routes and does not select optimal routes or prevent routing loops. Therefore, EGP cannot meet network management requirements.

BGP is another external gateway protocol designed to replace the original EGP. Unlike the original EGP, the BGP is able to perform routing optimization, avoid routing loops, transmit routes more efficiently, and maintain large amounts of routing information.

Although BGP is used to transmit routing information between ASs, it is not the best choice in some scenarios. For example, on the egress connecting a data center to the Internet, static routes instead of BGP are used for communication with external

networks to prevent a huge number of Internet routes from affecting the data center internal network.

BGP ensures high network security, flexibility, stability, reliability, and efficiency:

- BGP uses authentication and Generalized TTL Security Mechanism (GTSM) to ensure network security.
- BGP provides routing policies to allow for flexible route selection.
- BGP provides Route Summarization and Route Dampening to prevent route flapping and improve network stability.
- BGP uses the Transport Control Protocol (TCP) with port number 179 as the transport layer protocol and supports BFD for BGP, BGP Tracking, and BGP GR to improve network reliability.

18.2 Principles

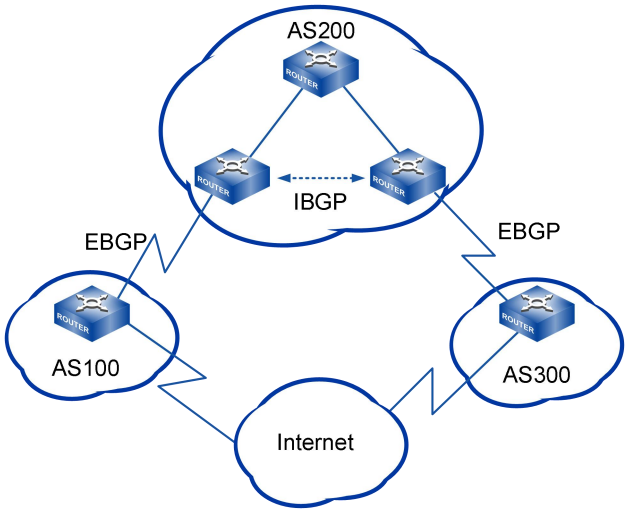
18.2.1 BGP Concepts

18.2.1.1 Autonomous System

An Autonomous System (AS) is a group of Internet Protocol (IP) networks that are controlled by one entity, typically an Internet service provider (ISP), and have the same routing policy. Each AS is assigned a unique AS number, which identifies an AS on a BGP network. Two types of AS numbers are available: 2-byte AS numbers and 4-byte AS numbers. A 2-byte AS number ranges from 1 to 65535, and a 4-byte AS number ranges from 1 to 4294967295. Devices supporting 4-byte AS numbers are compatible with devices supporting 2-byte AS numbers.

18.2.1.2 BGP Classification

As shown in the figure below, the BGP is divided by the way it operates into EBGP (External/Exterior BGP) and IBGP (Internal/Interior BGP).



- EBGP: EBGP runs between ASs. To prevent routing loops between ASs, a BGP device discards the routes with the local AS number when receiving the routes from EBGP peers.
- IBGP: IBGP runs within an AS. To prevent routing loops within an AS, a BGP device does not advertise the routes learned from an IBGP peer to the other IBGP peers and establishes full-mesh connections with all IBGP peers. To address the problem of too many IBGP connections between IBGP peers, BGP uses Route Reflector and BGP Confederation.

Note

If a BGP device needs to advertise the route received from an EBGP peer outside an AS through another BGP device, IBGP is recommended.

18.2.1.3 Device Roles in BGP Message Exchange

There are two device roles in BGP message exchange:

- Speaker The device that sends BGP messages is a BGP speaker. A BGP speaker receives and generates new routes, and advertises the routes to other BGP speakers.
- Peer BGP speakers that exchange messages with each other are BGP peers. A group of BGP peers sharing the same policies can form a peer group.

18.2.1.4 BGP Router ID

The BGP router ID is a 32-bit value that is often represented by an IPv4 address to identify a BGP device. It is carried in the Open message sent during the establishment

of a BGP session. When two BGP peers need to establish a BGP session, they each require a unique router ID. Otherwise, the two peers cannot establish a BGP session.

18.2.2 BGP Working Principles

BGP peer establishment, update, and deletion involve five types of messages, six state machine states, and five route exchange rules.

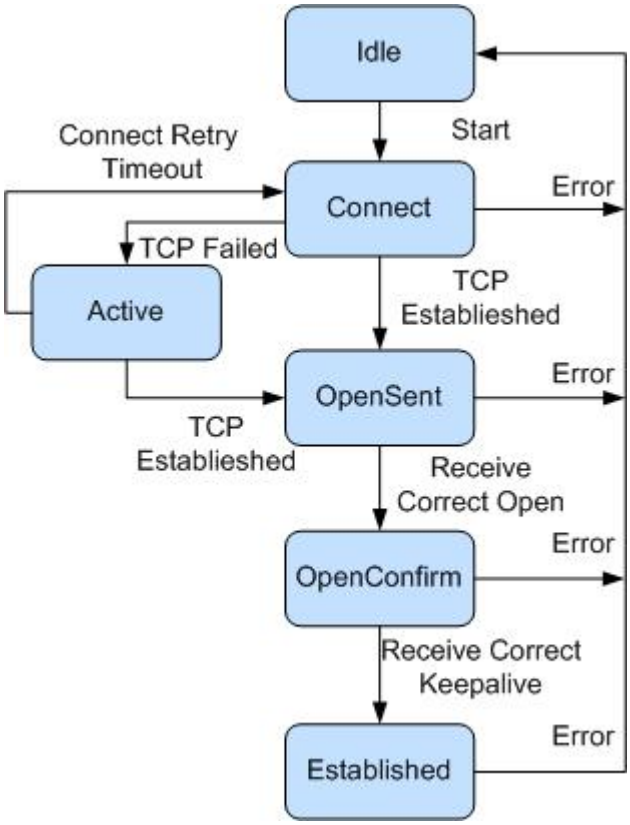
18.2.2.1 BGP Messages

BGP peers exchange the following messages, among which Keepalive messages are periodically sent and other messages are triggered by events.

- Open message: Used to establish BGP peer relationships.
- Update message: Used to exchange routes between BGP peers.
- Notification message: Used to terminate BGP connections.
- Keepalive message: Used to maintain BGP connections.
- Route-refresh message: Used to request the peer to retransmit routes if routing policies are changed. Only the BGP devices supporting route-refresh can send and respond to Route-refresh messages.

18.2.2.2 BGP State Machine

As shown in the figure below, there are 6 state machines in the interaction of BGP peers: Idle, Connect, Active, OpenSent, OpenConfirmed and Established. During the establishment of a BGP peer, the three states that are commonly visible are: Idle, Active and Established.



1 Idle is the initial state of the BGP. In the Idle state, the BGP rejects connection requests from neighbors. The BGP device initiates a TCP connection with its BGP peer and changes its state to Connect only after receiving a Start event from the system.

Notes

- The Start event occurs when an operator configures a BGP process or resets an existing BGP process or when the router software resets a BGP process.
- If an error occurs at any state of the FSM, for example, the BGP device receives a Notification packet or TCP connection termination notification, the BGP device returns to the Idle state.

2 In Connect state, the BGP device starts the Connect Retry timer and waits to establish a TCP connection.

- If the TCP connection is established, the BGP device sends an Open message to the peer and changes to the OpenSent state.
- If the TCP connection fails to be established, the BGP device moves to the Active state.
- If the BGP device does not receive a response from the peer before the Connect Retry timer expires, the BGP device attempts to establish a TCP connection with another peer and stays in Connect state.

3 In Active state, the BGP device keeps trying to establish a TCP connection with

the peer.

- If the TCP connection is established, the BGP device sends an Open message to the peer, closes the Connect Retry timer, and changes to the OpenSent state.
- If the TCP connection fails to be established, the BGP device stays in the Active state.
- If the BGP device does not receive a response from the peer before the Connect Retry timer expires, the BGP device returns to the Connect state.

4 In OpenSent state, the BGP device waits for an Open message from the peer and then checks the validity of the received Open message, including the AS number, version, and authentication password.

- If the received Open message is valid, the BGP device sends a Keepalive message and changes to the OpenConfirm state.
- If the received Open message is invalid, the BGP device sends a Notification message to the peer and returns to the Idle state.

5 In OpenConfirm state, the BGP device waits for a Keepalive or Notification message from the peer. If the BGP device receives a Keepalive message, it transitions to the Established state. If it receives a Notification message, it returns to the Idle state.

6 In Established state, the BGP device exchanges Update, Keepalive, Route-refresh, and Notification messages with the peer.

- If the BGP device receives a valid Update or Keepalive message, it considers that the peer is working properly and maintains the BGP connection with the peer.
- If the BGP device receives an invalid Update or Keepalive message, it sends a Notification message to the peer and returns to the Idle state.
- If the BGP device receives a Route-refresh message, it does not change its status.
- If the BGP device receives a Notification message, it returns to the Idle state.
- If the BGP device receives a TCP connection termination notification, it terminates the TCP connection with the peer and returns to the Idle state.

18.2.2.3 Route Exchange Rules

A BGP device adds optimal routes to the BGP routing table to generate BGP routes. After establishing a BGP peer relationship with a neighbor, the BGP device follows the following rules to exchange routes with the peer:

- Advertises the BGP routes received from IBGP peers only to its EBGP peers.
- Advertises the BGP routes received from EBGP peers to its EBGP peers and IBGP peers.
- Advertises the optimal route to its peers when there are multiple valid routes to the same destination.
- Sends only updated BGP routes when BGP routes change.
- Accepts all the routes sent from its peers.

18.2.3 Interaction Between BGP and IGPs

BGP and IGPs use different routing tables. To allow devices in different ASs to communicate, BGP needs to interact with IGPs. That is, BGP and IGP routes can be imported into IGP and BGP routing tables respectively.

18.2.3.1 Importing IGP Routes into BGP

BGP cannot discover routes and needs to import IGP routes into a BGP routing table for inter-AS communication. When an AS needs to advertise routes to another AS, an AS boundary router (ASBR) will import IGP routes into its BGP routing table. For better network planning, BGP can filter and set attributes for imported IGP routes based on routing policies. Alternatively, EBGP peers can select routes when traffic enters an AS based on the configured Multi-Exit Discriminator (MED) value.

BGP can import routes in either import or network mode:

- In import mode, BGP imports routes according to protocol types, such as RIP routes, OSPF routes, and IS-IS routes. To ensure validity of imported IGP routes, BGP can also import static routes and direct routes in import mode.
- In network mode, BGP imports routes in an IP routing table one by one. Therefore, the network mode is more precise than the import mode.

18.2.3.2 Importing BGP Routes into IGPs

When an AS needs to import routes from other ASs, an ASBR of this AS will import BGP routes into its IGP routing table. To prevent a large number of BGP routes from

affecting devices in this AS, IGPs can filter and set attributes for imported BGP routes based on routing policies.

18.2.3.3 Problem and Solution Related to Route Import Between BGP and IGPs

BGP and IGPs often import routes from each other to advertise routes between ASs. However, the following problems may occur:

- If a large number of BGP routes exist, low-end devices within ASs may be unable to have these routes installed, resulting in route loss.
- If a route is unstable, for example, a port frequently alternates between Up and Down states, all routes within an AS may flap, affecting network stability.
- BGP prevents routing loops using route attributes, for example, the AS_Path attribute. When all BGP routes are re-advertised into IGPs, their route attributes will be lost. As a result, the BGP routing loop prevention mechanism becomes ineffective, and routing loops may occur.

On a large IP network, the number of BGP routes is much larger than the number of IGP routes. Therefore, when BGP routes need to be imported into IGPs, exercise caution to prevent a large number of imported BGP routes from affecting IGP routes. To do so, use default routes and summarize routes to reduce the number of BGP routes.

18.2.4 BGP Route Selection Rules and Load Balancing

There may be multiple routes to the same destination in a BGP routing table. BGP will select one route as the optimal route and advertise it to peers. To select the optimal route among these routes, BGP compares the BGP attributes of the routes in sequence based on route selection rules.

18.2.4.1 BGP Attributes

Routing attributes are specific descriptions of routing. All BGP routing attributes can be classified into the following four categories. Common BGP attribute types are shown in the table below:

- Well-known mandatory attribute: All BGP devices can identify this type of attribute, which must be carried in Update messages. Without this type of attribute, errors occur in routing information.
- Well-known discretionary attribute: All BGP devices can identify this type of

attribute, which is optional in Update messages. Errors do not occur in routing information even if there is a lack of this type of attribute.

- Optional transitive attribute: A BGP device may not identify this type of attribute but still accepts and advertises them to peers.
- Optional non-transitive attribute: A BGP device may not identify this type of attribute, in which case it ignores them without advertising them to peers.

Attribute name	Type
Origin	Acceptance must be followed
AS_Path	Acceptance must be followed
Next_Hop	Acceptance must be followed
Local_Pref	Any recognized
MED	Optional non-transitive
Community	Optional transitive
Originator_ID Attribute	Optional non-transitive
Cluster_List Attribute	Optional non-transitive

The following describes common BGP route attributes:

- Origin

  The Origin attribute defines the origin of a route and marks the path of a BGP route. The Origin attribute is classified into three types:

  - IGP: A route with IGP as the Origin attribute has the highest priority. The Origin attribute of the routes imported into a BGP routing table using the network command is IGP.
  - EGP: A route with EGP as the Origin attribute has the second highest priority. The Origin attribute of the routes obtained through EGP is EGP.
  - Incomplete: A route with Incomplete as the Origin attribute has the lowest priority. The Origin attribute of the routes learned by other means is Incomplete. For example, the Origin attribute of the routes imported by BGP using the import-route command is Incomplete.

- AS_Path

  The AS_Path attribute records all of the ASs that a route passes through from the source to the destination in the vector order. To prevent inter-AS routing loops, a BGP device does not receive the routes whose AS_Path list contains the local AS number.

When a BGP speaker advertises an imported route:

- If the route is advertised to EBGP peers, the BGP speaker creates an AS_Path list containing the local AS number in an Update message.
- If the route is advertised to IBGP peers, the BGP speaker creates an empty AS_Path list in an Update message.

When a BGP speaker advertises a route learned in the Update message sent by another BGP speaker:

- If the route is advertised to EBGP peers, the BGP speaker adds the local AS number to the leftmost of the AS_Path list. According to the AS_Path list, the BGP speaker that receives the route can determine the ASs through which the route passes to reach the destination. The number of the AS that is nearest to the local AS is placed on the top of the AS_Path list. The other AS numbers are listed according to the sequence in which the route passes through ASs.
- If the route is advertised to IBGP peers, the BGP speaker does not change the AS_Path attribute of the route.

- Next_Hop

  The Next_Hop attribute records the next hop that a route passes through. The Next_Hop attribute of BGP is different from that of an IGP because it may not be a neighboring IP address. A BGP speaker processes the Next_Hop attribute based on the following rules:

  - When advertising a route to an EBGP peer, a BGP speaker sets the Next_Hop attribute of the route to the address of the local interface through which the BGP peer relationship is established.
  - When advertising a locally originated route to an IBGP peer, the BGP speaker sets the Next_Hop attribute of the route to the address of the local interface through which the BGP peer relationship is established.
  - When advertising a route learned from an EBGP peer to an IBGP peer, the BGP speaker does not change the Next_Hop attribute of the route.

- Local_Pref

  The Local_Pref attribute indicates the BGP preference of a device and helps determine the optimal route when traffic leaves an AS. When a BGP device obtains multiple routes to the same destination address but with different next hops from different IBGP peers, the BGP device prefers the route with the highest Local_Pref. The Local_Pref attribute is exchanged only between IBGP peers and is not advertised to other ASs. The Local_Pref attribute can be manually configured. If no Local_Pref attribute is configured for a route, the

Local_Pref attribute of the route uses the default value 100.

- MED

  The multi-exit discriminator (MED) attribute helps determine the optimal route when traffic enters an AS. When a BGP device obtains multiple routes to the same destination address but with different next hops from EBGP peers, the BGP device selects the route with the smallest MED value as the optimal route. The MED attribute is exchanged only between two neighboring ASs. The AS that receives the MED attribute does not advertise it to any other ASs. The MED attribute can be manually configured. If no MED attribute is configured for a route, the MED attribute of the route uses the default value 0.

- Community

  The Community attribute: Identifies the BGP routes with the same characteristics. Simplifies the applications of routing policies. Facilitates route maintenance and management.

  The Community attribute includes self-defined community attributes and well-known community attributes. Recognized group attributes are shown in the following table.

Community Attribute	Value	Note
Internet	0 (0x00000000)	A BGP device can advertise the received route with the Internet attribute to all peers.
No_Advertise	4294967042 (0xFFFFFF02)	A BGP device does not advertise the received route with the No_Advertise attribute to any peer.
No_Export	4294967041 (0xFFFFFF01)	A BGP device does not advertise the received route with the No_Export attribute to devices outside the local AS.
No_Export_Subconfed	4294967043 (0xFFFFFF03)	A BGP device does not advertise the received route with the No_Export_Subconfed attribute to devices outside the local AS or to devices outside the local sub-AS.

- **Originator_ID and Cluster_List**
  The Originator_ID attribute and Cluster_List attribute help eliminate loops in route reflector scenarios. For details, see Route Reflector.

18.2.4.2 BGP Route Selection Policies

When there are multiple routes to the same destination, BGP compares the following attributes in sequence to select the optimal route:

1 Prefers the route with the largest PrefVal value.
The protocol preference value (PrefVal) is a device-specific property that is only valid locally.

2 Prefers the route with the highest Local_Pref.
If a route does not have the Local_Pref attribute, the Local_Pref attribute of the route uses the default value 100.

3 Manual aggregation route, automatic aggregation route, route introduced by network command, route introduced by import-route command, and route learned from peers are preferred.

4 Prefers the route with the shortest AS_Path.

5 The route of Origin type is IGP, EGP, Incomplete.

6 Prefers the route with the lowest MED if routes are received from the same AS.

7 Prefers EBGP routes, IBGP routes, LocalCross routes, and RemoteCross routes.
LocalCross allows a PE to add the VPNv4 route of a VPN instance to the routing table of the VPN instance if the export RT of the VPNv4 route matches the import RT of another VPN instance on the PE. RemoteCross allows a local PE to add the VPNv4 route learned from a remote PE to the routing table of a VPN instance on the local PE if the export RT of the VPNv4 route matches the import RT of the VPN instance.

8 Prefers the route with the lowest IGP metric to the BGP next hop.

Notes
If there are multiple routes to the same destination, an IGP calculates the route metric using its routing algorithm.

9 Prefers the route with the shortest Cluster_List.

Notes
By default, Cluster_List takes precedence over Originator_ID during BGP route selection.

10 Prefers the route advertised by the device with the smallest router ID.
If a route carries the Originator_ID attribute, BGP prefers the route with the smallest Originator_ID without comparing the router ID.

11 Prefers the route learned from the peer with the lowest IP address.

18.2.4.3 BGP Load Balancing

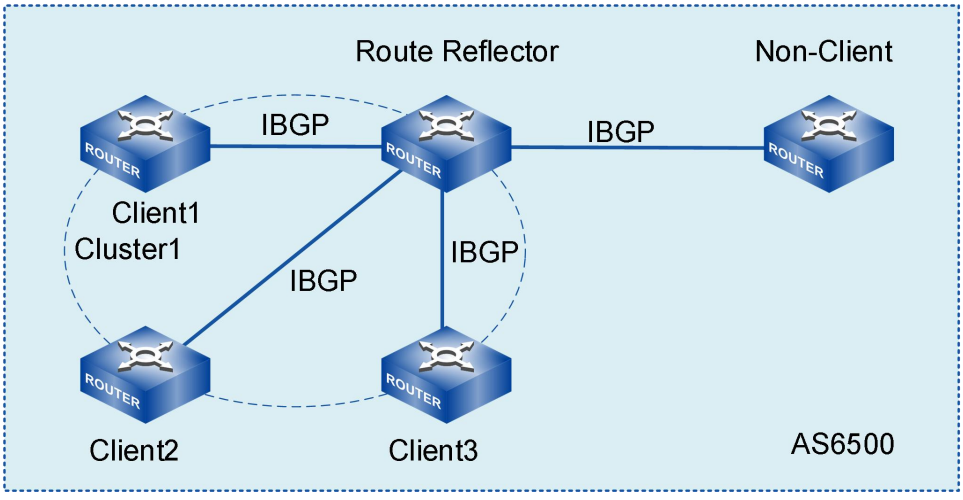
When there are multiple equal-cost routes to the same destination, you can perform load balancing among these routes to load balance traffic. Equal-cost BGP routes can be generated for traffic load balancing only when the first eight route attributes described in "BGP Route Selection Policies" are the same.

18.2.5 Route Reflection

To ensure connectivity between IBGP peers, you need to establish full-mesh connections between IBGP peers. If there are n devices in an AS, n(n-1)/2 IBGP connections need to be established. When there are a large number of devices, many network resources and CPU resources are consumed. A route reflector (RR) can be used between IBGP peers to solve this problem.

18.2.5.1 Roles in RR

AS shown below, there are several roles within an AS regarding routing reflectors:



- Route reflector (RR): A BGP device that can reflect the routes learned from an IBGP peer to other IBGP peers. An RR is similar to a designated router (DR) on an OSPF network.

- Client: An IBGP device whose routes are reflected by the RR to other IBGP devices. In an AS, clients need to connect only to the RR directly.
- Non-client: An IBGP device that is neither an RR nor a client. In an AS, a non-client must establish full-mesh connections with the RR and all the other non-clients.
- Originator: A device that originates routes in an AS. The Originator_ID attribute helps eliminate routing loops in a cluster.
- Cluster: A set of the RR and clients. The Cluster_List attribute helps eliminate routing loops between clusters.

18.2.5.2 RR Principles

Clients in a cluster need to exchange routing information only with the RR in the same cluster. Therefore, clients need to establish IBGP connections only with the RR. This reduces the number of IBGP connections in the cluster. As shown above, within the AS65000 one device ACTS as RR and three devices act as clients to form Cluster1. The number of IBGP connections in AS 65000 is then reduced from 10 to 4, which simplifies the device configuration and reduces the loads on the network and CPU.

RR goes beyond "BGP routing from AN IBGP peer, where the BGP device is only published to its EBGP peer." The unique Cluster_List attribute and Originator_ID attribute are used to prevent routing loops. The unique Cluster_List attribute and Originator_ID attribute are used to prevent routing loops. The RR advertises routes to IBGP peers based on the following rules:

- The RR advertises the routes learned from a non-client to all the clients.
- The RR advertises the routes learned from a client to all the other clients and all the non-clients.
- The RR advertises the routes learned from an EBGP peer to all the clients and non-clients.

18.2.5.3 Cluster_List Attribute

An RR and its clients form a cluster, which is identified by a unique cluster ID in an AS. To prevent routing loops between clusters, an RR uses the Cluster_List attribute to record the cluster IDs of all the clusters that a route passes through based on the following rules:

- When a route is reflected by an RR for the first time, the RR adds the local cluster ID to the top of the cluster list. If there is no cluster list, the RR creates a Cluster_List attribute.
- When receiving an updated route, the RR checks the cluster list of the route. If the cluster list contains the local cluster ID, the RR discards the route. If the

cluster list does not contain the local cluster ID, the RR adds the local cluster ID to the cluster list and then reflects the route.

18.2.5.4 Originator_ID Attribute

An originator ID identifies the originator of a route and is generated by an RR to prevent routing loops in a cluster. Its value is the same as a router ID.

- When a route is reflected by an RR for the first time, the RR adds the Originator_ID attribute to this route. The Originator_ID attribute identifies the originator of the route. If the route contains the Originator_ID attribute, the RR retains this Originator_ID attribute.
- When a device receives a route, the device compares the originator ID of the route with the local router ID. If they are the same, the device discards the route.

18.2.5.5 Backup RR

To ensure network reliability and prevent single-point failures, a cluster requires redundant RRs. An RR allows a BGP device to advertise the routes received from an IBGP peer to other IBGP peers. Therefore, routing loops may occur between RRs in the same cluster. To prevent this, all the RRs in the cluster must use the same cluster ID.

18.2.5.6 RRs of Multiple Clusters in an AS

If there are multiple clusters in an AS, RRs of the clusters establish IBGP peer relationships. When RRs reside at different network layers, an RR at the lower network layer can be configured as a client to implement hierarchical RR. When RRs reside at the same network layer, RRs of different clusters can establish full-mesh connections to implement flat RR.

18.2.6 BGP Confederation

In addition to a route reflector, the confederation is another method that reduces the number of IBGP connections in an AS. A confederation divides an AS into sub-ASs. Full-mesh IBGP connections are established in each sub-AS. EBGP connections are established between sub-ASs. ASs outside a confederation still consider the confederation as an AS. After a confederation divides an AS into sub-ASs, it assigns a

confederation ID (the AS number) to each router within the AS. This brings two benefits. Original IBGP attributes are retained, including the Local_Pref attribute, MED attribute, and Next_Hop attribute. Confederation-related attributes are automatically deleted when being advertised outside a confederation. Therefore, the administrator does not need to configure the rules for filtering information such as sub-AS numbers at the egress of a confederation.

Comparisons Between a Route Reflector and a Confederation

Route Reflection	Confederation
There is no need to change the existing network topology. Compatibility is good.	Logical topology needed to change.
It is easy to configure; you just need to configure the device as a reflector, and the client does not need to know that it is a client.	All devices need to be reconfigured.
Full connections between clusters are still required.	There are special EBGP connections between the subAs of the alliance, and full connections are not required.
Suitable for medium and large scale network.	Suitable for large-scale network.

18.2.7 Route Summarization

The BGP routing table of each device is large on large networks. This burdens devices, increases the route flapping probability, and affects network stability.

Route summarization is a mechanism that combines multiple routes into one route. This mechanism allows a BGP device to advertise only the summarized route but not all the specific routes to peers, thereby reducing the size of the BGP routing table. If the summarized route flaps, the network is not affected, improving network stability.

BGP supports automatic summarization and manual summarization on IPv4 networks, and supports only manual summarization on IPv6 networks. Automatic and manual summarization work as follows:

- Automatic summarization: Summarizes the routes imported by BGP. After automatic summarization is configured, BGP summarizes routes based on the

natural network segment and advertises only the summarized route to peers. For example, BGP summarizes 10.1.1.1/24 and 10.2.1.1/24 (two Class A addresses with non-natural mask) into 10.0.0.0/8 (Class A address with natural mask).

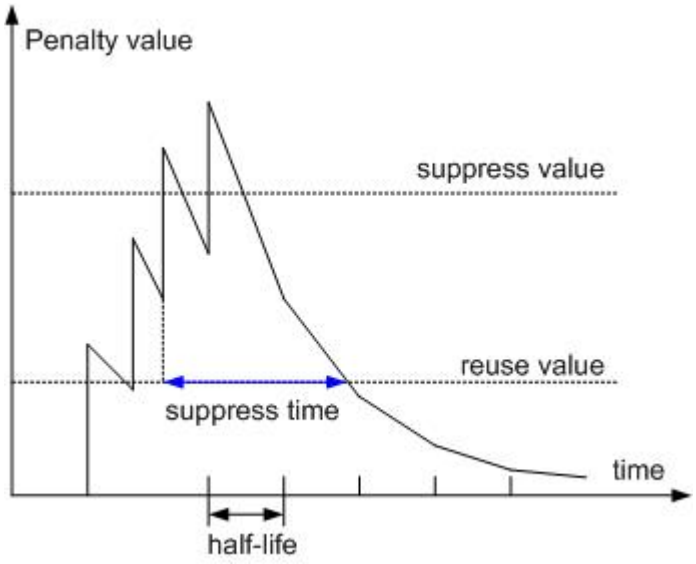
- Manual summarization: Summarizes routes in the local BGP routing table. Manual summarization can help control the attributes of the summarized route and determine whether to advertise specific routes.

To prevent routing loops caused by route summarization, BGP uses the AS_Set attribute. The AS_Set attribute is an unordered set of all ASs that a route passes through. When the summarized route enters an AS in the AS_Set attribute again, BGP finds that the local AS number has been recorded in the AS_Set attribute of the route and discards this route to prevent a routing loop.

18.2.8 Route Dampening

When BGP is used on complex networks, route flapping occurs frequently. To prevent this, BGP uses route dampening to suppress unstable routes.

Route flapping is a process of adding a route to an IP routing table and then withdrawing this route. When route flapping occurs, a BGP device sends an Update message to its neighbors. The devices that receive the Update message need to recalculate routes and modify routing tables. Frequent route flapping consumes lots of bandwidths and CPU resources and even affects normal network operation.



Route dampening measures the stability of a route using a penalty value. A larger penalty value indicates a less stable route. As shown in the figure above, each time the route oscillates, the BGP adds a penalty of 1000 to the route, and the rest of the time the penalty drops slowly. When the penalty value of a route exceeds the

suppression threshold, BGP suppresses this route, and does not add it to the IP routing table or advertise any Update message to peers. After a route is suppressed for a period of time (half-life), the penalty value is reduced by half. When the penalty value of a route decreases to the reuse threshold, the route is reusable and is added to the routing table. At the same time, BGP advertises an Update message to peers. The suppression time is the period from when a route is suppressed to when the route is reusable.

Route dampening applies only to EBGP routes but not IBGP routes. IBGP routes may include the routes of the local AS, and an IGP network requires that the routing tables of devices within an AS be the same. If IBGP routes are dampened, routing tables on devices are inconsistent when these devices have different dampening parameters. Therefore, route dampening does not apply to IBGP routes.

18.3 BGP Configuration

18.3.1 Enter BGP Route Configuration View

【Command】

```
router bgp < AS-NUMBER >
no router bgp < AS-number >
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

As-number: autonomous system number and the range is 1-4294967295.

【Description】

router bgp: the command is used to start the BGP process and enter BGP process mode.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
```

18.3.2 Enter the IPv4 or IPv6 Address Family

【Command】

address-family { ipv4 | ipv6}

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

address-family: the command is used to enter the ipv4/ ipv6 address family to configure BGP configuration mode.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router)#address-family ipv4
Switch(config-router-af)#
```

18.3.3 BGP Route Aggregation Address

【Command】

aggregate-address <A.B.C.D/M> **[as-set] [summary-only]**
no aggregate-address <A.B.C.D/M> **[as-set] [summary-only]**

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D/M: network address and mask.

as-set: generates a route with AS collection segments.

Summary-only: only the aggregated routes are announced, all routes are announced by default.

【Description】

aggregate-address: the command is used to create an aggregate route in the BGP routing table.

By default, routing aggregation is not performed.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #aggregate-address 10.0.0.0 255.0.0.0 as-set
```

18.3.4 BGP Always Comparison MED Value

【Command】

bgp always-compare-med
no bgp always-compare-med

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp always-compare-med: the command is used to set the BGP always compare MED.

By default, MED is only compared when AS is the same.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp always-compare-med
```

18.3.5 The BGP optimal Path Ignore the AS-Path Length

【Command】

bgp bestpath as-path ignore

no bgp bestpath as-path ignore

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp bestpath as-path ignore: the command is used to configure to ignore the as-path length.

By default, the as-path length is considered when selecting the optimal path.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp bestpath as-path ignore
```

18.3.6 Comparison of AS-Path Length of BGP Optimum Path

【Command】

bgp bestpath compare-confed-aspath

no bgp bestpath compare-confed-aspath

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp bestpath compare-confed-aspath: the command is used to allow the same external routing to compare the aspath path length of the federation, the smaller the aspath length within the federation, the higher the path priority.

By default, the optimal path is chosen not by comparing ASPATHS from EBGP peer routes within the same federation, routing is optimized according to other conditions.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp bestpath compare-confed-aspath
```

18.3.7 BGP Optimum Path Routing ID Comparison

【Command】

```
bgp bestpath compare-routerid
no bgp bestpath compare-routerid
```

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp bestpath compare-routerid: the command is used for router ID that allows the same external routing comparison path, the smaller the router ID, the higher the path priority.

By default, when selecting the optimal path, two paths with the same path attributes will be received by different EBGP peers, The first one is not the one with the highest priority.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp bestpath compare-routerid
```

18.3.8 MED Property of BGP Optimum Path

【Command】

```
bgp bestpath med {[confed] [missing-as-worst]}
```

no bgp bestpath med {[confed] [missing-as-worst]}

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

Confed: compare MED in alliance path.

missing-as-worst: consider that the value of MED is the greatest when it is missing.

【Description】

bgp bestpath med confed: the command is used to compare the properties of med in the alliance path and to consider its value to be the maximum when the med is missing.

By default, the MED value of the path of the peers is not compared when selecting the optimal path.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp bestpath med confed
```

18.3.9 BGP Route Reflection

【Command】

bgp client-to-client reflection

no bgp client-to-client reflection

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp client-to-client reflection: the command is used to configure whether to process route reflection.

By default, route reflection is performed when the client is configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp client-to-client reflection
```

18.3.10 Configure the Group Identifier for the Routing Reflector

【Command】

```
bgp cluster-id {(A.B.C.D) | <0-4294967295>}
no bgp cluster-id
```

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: group identification, IP address format.

<0-4294967295> : group identification, decimal integer, value range is 0 ~ 4294967295, the system will process it into IP address format.

【Description】

bgp cluster-id: the command is used to configure the group identification of the route reflector.

By default, the identification is router-id of the router reflector.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp cluster-id 1.1.1.1
```

18.3.11 Configure the Identifier for the AS Alliance

【Command】

```
bgp confederation identifier <NUMBER>
no bgp confederation identifier [<NUMBER>]
```

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

Number: confederation identifier

【Description】

bgp confederation identifier: the command is used to configure the identifier for the AS confederation.

By default confederation identifier is not configured.

Confederation is another way to reduce the number of IBGP peer connections in autonomous systems. An autonomous system is divided into multiple sub-autonomous systems, and these sub-autonomous systems are formed into a confederation by setting a unified federation ID(i.e., confederation AS number). For the outside of the confederation, the whole confederation is still regarded as an AS, and only the AS number of the confederation is visible to the external. Within the confederation, full IBGP peer-to-peer connections are still established between BGP speakers within the sub-autonomous systems and EBGP connections between BGP speakers in the sub-autonomous systems. Although EBGP links have been established between BGP speakers in sub-autonomous systems, information exchange remains the same for next hop, MED, local priority, etc.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp confederation identifier 65000
```

18.3.12 Configure the Member AS for the AS Alliance

【Command】

```
bgp confederation peers <as-number> [<as-number>]
no bgp confederation peers <as-number> [<as-number>]
```

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

as-number: AS, a member of the confederation. range :1-65535.

【Description】

bgp confederation peers: the command is used to configure the identifier for the AS confederation.

By default, confederation member is not configured.

Confederation is a way to reduce the number of IBGP peer connections in autonomous systems. An autonomous system is divided into multiple sub-autonomous systems, and these sub-autonomous systems are formed into a confederation by setting a unified federation ID(i.e., confederation AS number). For the outside of the confederation, the whole confederation is still regarded as an AS, and there are other AS numbers of the confederation are visible to the external. Within the confederation, full IBGP peer-to-peer connections are still established between BGP speakers within the sub-autonomous systems and EBGP connections between BGP speakers in the sub-autonomous systems. Although EBGP links have been established between BGP speakers in sub-autonomous systems, information exchange remains the same for next hop, MED, local priority, etc. This command specifies AS, a member of the alliance.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp confederation peers 100 110
```

18.3.13 EBGP Route Suppression

【Command】

bgp dampening {[<half-life-reach>] [<reusing> <suppressing> <duration>] [<half-life-unreach>] | [route-map <WORD>] }

no bgp dampening {[<half-life>] [<reusing> <suppressing> <duration>] [<half-life-unreach>] | [route-map [<WORD>]] }

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

half-life-reach: specify the half-life of reachable routes, specify the interval between which the penalty value is halved, unit minutes, range 1-45 minutes, default is 15 minutes.

Reusing: do not dampen the routing when the penalty value drops to this value, ranging from 1-20000, default to 750.

Suppressing: dampen the routing when the penalty value reaches the value, the range is 1-20000, the default value is 2000.

duration: the maximum duration of routing suppression, beyond which the routing suppression is automatically unmounted, is in minutes, ranging from 1 to 255, default is 60 minutes.

half-life-unreach: Specify the half-life of the unreachable route, specify how long the interval is, and halve the penalty value, unit minute, range 1-45 minutes, default is 15 minutes.

WORD: the name of oute-map, which applies routing attenuation to specific routes via route-map.

Route attenuation is applied to all routes by default.

【Description】

`bgp dampening`: the command is used to suppress unstable EBGP route and is invalid to the IBGP routing.

BGP uses the concept of penalty values to describe the stability of the route. The higher the penalty value, the more unstable the route. The penalty value is increased by 1000 per routing concussion (when withdraw message is received). When the penalty value increases to a certain level, it will not increase again. This value is called the penalty upper limit. This value depends on the user-configured duration, calculated by penalizing the upper limit = 2 ^ (duration/half-life) * reusing. At the same time, the upper limit of penalty value cannot be greater than 20000, so duration, half-life and reusing values should be adjusted according to the network conditions during configuration. These parameters roughly satisfy the following requires: 1) half-time and duration: half-time ≤ duration; 2) reusing, suppressing and penalty uper limit: reusing ≤ suppressing ≤ penalty uper limit. The user can also specify other half - life value, the value of duration is (half - life * 4), and reusing and suppressing values are 750 and 2000 respectively. EBGP routes with penalty value exceeds suppressing value will be suppressed, suppressed routing in BGP routing in the

process of the election will not be used, also won't notice to other BGP peers. If the suppressed route continues concussion, the penalty will continue to rise to the penalty upper limit. For every half-life time a suppressed route passes, the penalty is reduced by half. When the penalty value is reduced to the reusing value, the route that was updated at the last update will re-participate in the BGP routing election. When the penalty value is reduced to 0, the route that is withdraw message at the last update will be deleted from the BGP routing table.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp dampening 30 1200 10000 120
```

18.3.14 Set the BGP Address Family to IPv4 Unicast by Default

【Command】

```
bgp default ipv4-unicast
no bgp default ipv4-unicast
```

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp default ipv4-unicast: the command is used to set the address family to a unicast address that defaults to ipv4.

By default, IPv4 unicast is used for neighbor relationship build.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp default ipv4-unicast
```

18.3.15 The Default Local Priority Value for BGP

【Command】

bgp default local-preference ***<0-4294967295>***

no bgp default local-preference [*<0-4294967295>*]

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

<0-4294967295>: default local priority.

【Description】

bgp default local-preference: the command is used to set the default local-preference property value.

By default, the local priority value is 100.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp default local-preference 200
```

18.3.16 Compare MED Values of the Same AS Counterparts

【Command】

bgp deterministic-med

no bgp deterministic-med

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp deterministic-med: the command is used to set path precedence for comparing med values from peers of the same AS.

By default, comparisons are made in path receive order, the most recently received path being compared first.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp deterministic-med
```

18.3.17 Force the AS-Path of the First Location to be the EBGP Route

【Command】

bgp enforce-first-as

no bgp enforce-first-as

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp enforce-first-as: the command is used to set up that when BGP receives an external route, the forced route's as-path must contain the neighbor's as number in the position of the first as or break the peer's connection.

By default, it is not configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp enforce-first-as
```

18.3.18 Reset the Direct Connection EBGP Fault Interface

【Command】

bgp fast-external-failover

no bgp fast-external-failover

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp fast-external-failover: the command is used to set up a BGP session connection to be quickly closed when the network interface used by the directly connected EBGP peer to establish the connection fails.

By default, this function is configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp fast-external-failover
```

18.3.19 Record BGP Status Change Information

【Command】

bgp log-neighbor-changes

no bgp log-neighbor-changes

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

bgp log-neighbor-changesr: the command is used to enable the state message of the BGP to be recorded when the debug is not turned on.

By default, this function is configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp log-neighbor-changes
```

18.3.20 BGP Route ID

【Command】

```
bgp router-id <A.B.C.D >
no bgp router-id [<A.B.C.D >]
```

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: router ID

【Description】

bgp router-id: the command is used to set the ID of the router.

By default, the router ID is automatically obtained.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp router-id 1.1.1.1
```

18.3.21 BGP Next Hop Detection Interval

【Command】

```
bgp scan-time <interval>
no bgp scan-time [<interval>]
```

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

interval: the interval between checks.

【Description】

bgp scan-time: the command is used to set the time interval between which BGP periodically checks whether the next hop of the route is valid.

By default, the interval is 60.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #bgp scan-time 45
```

18.3.22 Specifies the Management Distance for the Routing Prefix

【Command】

```
distance <distance> <A.B.C.D/M>
no distance <distance> <A.B.C.D/M>
```

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

distance: management distance, the range is 1-255.

A.B.C.D/M: route prefix.

【Description】

distance: the command is used to set the administrative distance for the specified route prefix.

By default, it is not configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #distance 100 10.12.1.0/24
```

18.3.23 Managed Distances for Internal/External/Local BGP Routes

【Command】

distance bgp <EXTERNAL> <INTERNAL> <LOCAL>

no distance bgp <EXTERNAL> <INTERNAL> <LOCAL>

【View】

BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

External: BGP external (EBGP) management distance, the range is 1-255.

Internal: BGP internal (IBGP) management distance, the range is 1-255.

local: local management distance, the range is 1-255.

【Description】

distance bgp: the command is used to set different management distances for different types of BGP routes.

By default, EBGP management distance is 20, IBGP management distance is 200, and local route management distance is 200.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #distance bgp 30 180 180
```

18.3.24 Exit the BGP IPv4 or IPv6 Address Family

【Command】

exit-address-family

【View】

IPv4 or IPv6 Address Family of BGP Configuration View

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

The exit-address-family: the command is used to exit the BGP IPv4 or IPv6 address family configuration view.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router) #address-family ipv4
Switch(config-router-af)#exit-address-family
```

18.3.25 Network Segment Routing Notification BGP Routing Table

【Command】

network {<A.B.C.D> [mask <MASK>] **| <A.B.C.D/M>} [route-map <*map-name*>] [backdoor]**

no network {<A.B.C.D> [mask <MASK>] **| <A.B.C.D/M>} [route-map <*map-name*>] [backdoor]**

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D/M: network address and mask.

<A.B.C.D><MASK>: network address and mask.

map-name: the name of the route-map. route map name should be no more than 32 characters.

backdoor: this route is backdoor route. Enables router to select the route learned from IGP first instead of the route learned from IBGP neighbors.

【Description】

network: the command is used to notify segment route to the BGP route table.

By default, BGP does not notify any network segment route.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#network 10.11.20.0/24
```

18.3.26 Route Synchronization Notify Network Information

【Command】

network synchronization

no network synchronization

【View】

BGP configuration view, the IPv4 or IPv6 address family configuration view of the BGP

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

network synchronization: the command is used to configure the synchronize of BGP and the local route before announcing the network information configured by the network.

By default, BGP directly announces the network route, regardless of whether it is synchronized with the local route.

It is not recommended to turn off the switch, which may cause a routing black hole.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router)#network synchronization

Switch > enable
Switch#configure terminal
```

```
Switch(config)#router bgp 100
Switch(config)#address-family ipv4
Switch(config-router-af)#network synchronization
```

18.3.27 Configure BGP Peers/Peer Groups

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} remote-as < AS-number >
no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} remote-as < AS-number >

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

As-number: BGP peer (group) autonomous system number and the range is 1-65535.

【Description】

neighbor remote-as: the command is used to configure the BGP peer/peer group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
```

18.3.28 Activate the Specified Peer/Peer Group

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} activate
no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} activate

【View】

BGP configuration view,IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor activate: the command is used to activate the specified peer/peer group.

By default, it is active.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 activate
```

18.3.29 The Time Interval Between Sending the BGP Route Update Message

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} advertisement-interval < seconds >

no neighbor {<A.B.C.D> | <X:X::X:X> | <WORD>} advertisement-interval < seconds >

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

seconds: the time interval in which routing updates are sent. Range: 0-600 seconds.

【Description】

neighbor advertisement-interval: the command is used to set the time interval between sending BGP routing updates.

By default, the same routing update message is sent to an internal peer group at a 5-second interval and to an external peer group at a 30-second interval.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 advertisement-interval 10
```

18.3.30 The Number of Occurrences of the Same AS in the Neighbor Route AS List

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} allowas-in [< numbers >]

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} allowas-in [< numbers>]

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

Numbers: the times an AS number is allowed to repeat. range :1-10.

【Description】

neighbor allowas-in: the command is used to specify the number of times the same AS is allowed in the neighbor's routing AS list.

By default, AS is not allowed to be repeated on the same route. When repetition is configured to allow , the default repeat times allowed without the <1-10> parameter is 3.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
```

```
Switch(config- router)#neighbor 1.1.1.1 allowas-in 10
```

18.3.31 The Time Interval Between Sending the Local Originating BGP Route

【Command】

```
neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} as-origination-interval
<seconds >
no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>}
as-origination-interval <seconds >
```

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

seconds: the time interval between sending local originating BGP routes. Range: 1-600 seconds.

【Description】

The neighbor as-origination interval command is used to configure the time interval between notifying a specified peer of a local originating BGP route.

By default, the time interval for sending a local originating BGP route is 15 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 as-origination-interval
100
```

18.3.32 The Routing Properties Associated with the Transport Neighbors Remain Unchanged

【Command】

```
neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} attribute-unchanged
[as-path] [med] [next-hop]
no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} attribute-unchanged
[as-path] [med] [next-hop]
```

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group

as-path: as-path property.

med: med property.

next-hop ： next-hop property

【Description】

The neighbor attribute-unchanged command is used to configure that the related routing property will not be changed when passed to the specified neighbor.

【Instance】

```
Switch> enable
Switch#configure terminal
Switc (config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 attribute-unchanged
```

18.3.33 Dynamic Updates and Routing Refreshes with Neighbors

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} capability {dynamic |route-refresh}

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} capability {dynamic |route-refresh}

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

Dynamic: dynamic update capability negotiation.

route-refresh: route refresh capability negotiation.

【Description】

neighbor capability: The command is used to configure dynamic update and routing refresh capability negotiation with neighbors. Through this configuration, negotiate the capability supported by both parties in the OPEN message. If the capability is supported by the end, a response will be given; otherwise, NOTIFICATION will be sent, and the initiator will send OPEN without the capability to re-establish the connection. Dynamic capability means that a connection does not have to be restarted when the supported address family negotiation changes. Routing refresh is to issue a routing refresh request when some properties can be configured for soft reset and the other party can resend the existing route to the originating end. With the routing refresh property configured, user can use the clear ip bgp * soft in command to refresh when changing policy without rebuilding the connection.

By default: configure routing refresh capability instead of dynamic update capability.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
```

```
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 capability dynamic
```

18.3.34 Egress Route Filter With Neighbor

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} capability orf prefix-list {<both>|<send>|<receive>}

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} capability orf prefix-list {<both>|<send>|<receive>}

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

both:receive and send the prefix- List filter rule.

send: sends the prefix-list filter rule.

receive: receives the prefix-list filter rule.

【Description】

neighbor capability orf prefix-list: the command is used to configure the outlet with the neighbor to be negotiated by the filter function. Through this configuration, negotiate the capability supported by both parties in the OPEN message. If the capability is supported by the end, a response will be given; otherwise, NOTIFICATION will be sent, and the initiator will send OPEN without the capability to re-establish the connection. After this capability is configured, the party that has the prefix-list filter rule configured sends its own filter rule to the peer when the connection is established, and the peer applies its own rule to avoid sending the route that will be filtered on the other side.

By default ORF function is not configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
```

```
Switch(config-router)#neighbor 1.1.1.1 remote-as 200
Switch(config-router)#neighbor 1.1.1.1 capability orf prefix-list
both
```

18.3.35 TCP Connection Collision Detection with Neighbor

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} collide-established

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} collide-established

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor collide-established: the command is used to enable conflict detection and resolution in the event of a TCP connection conflict.

By default, it is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router)#neighbor 1.1.1.1 remote-as 200
Switch(config-router)#neighbor 1.1.1.1 collide-established
```

18.3.36 Send a Default Route to a Peer/Peer Group

【Command】

neighbor {*<A.B.C.D >* | *<X:X::X:X>* | *<WORD>*} default-originate [route-map *<WORD>*]

no neighbor {*<A.B.C.D >* | *<X:X::X:X>* | *<WORD>*} default-originate [route-map *<WORD>*]

【View】

BGP configuration view,IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

WORD: the name of the route map.

【Description】

neighbor default-originate: the command is used to send the default route to the peer/peer group.

By default: no default route is sent to peer/peer groups.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 default-originate
```

18.3.37 Peer/Peer Group Description Information

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} description <LINE>
no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} description [< LINE >]

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

LINE: description information for peers.

【Description】

neighbor description: the command is used to configure description information for peer/peer groups.

By default: peer/peer group does not have description information.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 description test
```

18.3.38 Filtering Strategy Based on ACL

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} distribute-list {<access-list-number> | <WORD>} {in | out}

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} distribute-list {<access-list-number> | <WORD>} {in | out}

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

access-list-number: ACL number, value range 1-199 or 1300-2699.

WORD: ACL name.

【Description】

neighbor distribution-list: the command is used to set up acl-based filtering policies for peer and peer groups.

By default: no ACL based filtering policy is set for peer/peer groups.

Whether for the input (in) or output rules (out), at any time, the command and neighbor prefix-list can not coexist, only one command is effective.

【Instance】

```
Switch> enable
```

```
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 distribute-list 1 in
```

18.3.39 No Capacity Negotiation when Establishing a Connection

【Command】

neighbor {<A.B.C.D> | <X:X::X:X> | <WORD>} dont-capability-negotiate

no neighbor {<A.B.C.D> | <X:X::X:X> | <WORD>} dont-capability-negotiate

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor dont-capability-negotiate: the command is used to configure that do not carry out the ability negotiation when a connection is established.

By default, ability negotiation is carried out.

By default, capability negotiation occurs when the connection is established with the other side, so this configuration allows users to establish a connection without capability negotiation when knowing that the other side is an older version of BGP that does not support capability negotiation.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router)#neighbor 1.1.1.1 remote-as 200
Switch(config-router)#neighbor 1.1.1.1 dont-capability-negotiate
```

18.3.40 Allows Non-directly Connected Neighbors to Establish EBGP Sessions

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} ebgp-multihop [<count>]

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} ebgp-multihop [<count>]

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

Count: number of hops allowed, range 1-255.

【Description】

neighbor ebgp-multihop: the command is used to configure to allow EBGP sessions to be established with neighbors on non-directly connected networks.

By default: EBGP sessions are not allowed with neighbors on indirectly connected networks.

Set the parameter count to configure the maximum router hops for EBGP sessions at the same time, and 255 hops without the parameter count.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 ebgp-multihop
```

18.3.41 Force Direct Contact with Neighbors for Multiple Hop

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} enforce-multihop

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} enforce-multihop

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor enforce-multihop: the command is used to configure the connection to the neighbor to multihop.

By default, it is not configured.

Direct connection routing cannot be forced to be multi-hop, but with this configuration, the system can be forced to treat the connection as a multi-hop connection. That is to say, the checks that were originally made only for IBGP and EBGP of non-directly connected routes, all are checked after this property is set. At the ingress, there is no direct connection check for the next jump in case of forced multi-jump.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 enforce-multihop
```

18.3.42 AS-Path Access List Filtering

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} filter-list <WORD> {in | out}

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} filter-list <WORD> {in | out}

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

WORD: as-path list name.

【Description】

neighbor filter-list: the command is used to configure access list control for AS-PATH when routing information is sent and received with a specified BGP peer (group).

By default, it is not configured.

After the IP AS-PAT H access list is first configured and applied to the specified neighbor, user can control to send/receive route with the specified AS number in the AS list to that neighbor. Accept and reject depend on settings of the access list, while issue and receive are set by this command.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ip as-path access-list test deny 100
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 filter-list test out
```

18.3.43 Configure the Interface to the BGP Peer/Peer Group

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} interface <WORD>

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} interface <WORD>

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

WORD: interface name.

【Description】

neighbor interface: the command is used to configure the interface connected to the BGP peer (group).

By default, it is not configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 interface vlanif2
```

18.3.44 Number of Received Peer Prefixes

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} maximum-prefix <limit> [<threshold>] [warning-only]

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} maximum-prefix <limit> [<threshold>] [warning-only]

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

limit: the upper limit of the number of items allowed to receive routing information, ranging from 1-- -4294967295.

threshold: specifies a value that starts generation of warnings and a maximum number of percentages, ranging from 1 to 100, and defaults to 75.

warning-only: when routing information reaches the limit, the BGP connection is terminated and a log message is generated.

【Description】

neighbor maximum-prefix: the command is used to configure the limit on the number of prefixes received from the specified BGP peer.

By default, it is not limited.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 maximum-prefix 100
```

18.3.45 Force Next-hop Addresses as Neighbors

【Command】

```
neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} next-hop-self
no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} next-hop-self
```

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor next-hop-self: the command is used to configure a route that requires the neighbor to make the next hop to the local route.

By default, it is not configured.

In the EBGP environment, the next hop automatically points to the source neighbor, but in the IBGP environment, if the same network segment is routed, the original next hop remains unchanged. But if it is not a broadcast network, there will be a problem,

so this command can be used to force to use itself as neighbor's next hop as long as it is in IBGP.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 next-hop-self
```

18.3.46 Ignore the Results of Peer Capacity Negotiation

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} override-capability
no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} override-capability

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor override-capability: the command is used to configure the result of ignoring the BGP peer (group) capability negotiation.

By default, it is not configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 override-capability
```

18.3.47 The Peer Connection is Passive

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} passive

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} passive

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor passive: the command is used to configure the connection to the BGP peer (group) to be passive and not to actively send the connection request.

By default: initiate the connection.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 passive
```

18.3.48 Create a BGP Peer Group

【Command】

neighbor <WORD> peer-group

no neighbor <WORD> peer-group

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

WORD: name of the peer group

【Description】

neighbor <WORD>peer-group: the command is used to create the BGP peer group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor test peer-group
```

18.3.49 Add a Peer Group Member

【Command】

neighbor {<A.B.C.D > | <X:X::X:X>} peer-group <WORD>
no neighbor {<A.B.C.D > | <X:X::X:X> } peer-group <WORD>

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address

X:X::X:X: IPv6 peer interface address

WORD: name of the peer group

【Description】

neighbor {< A.B.C.D> | <X:X: X:X>} peer group <WORD> : the command is used to add a specified BGP peer to a peer group.

By default，it have not joined the peer group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor test peer-group
Switch(config- router)#neighbor 1.1.1.1 peer-group test
```

18.3.50 TCP Port Number for Communicating with Neighbors

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} port < number >

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} port < number >

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

number: TCP port number, the range is 0-65535.

【Description】

neighbor passive: the command is used to configure the TCP port number to communicate with that neighbor.

By default, the port number is 179.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router)#neighbor 1.1.1.1 remote-as 200
Switch(config-router)#neighbor 1.1.1.1 port 1022
```

18.3.51 Routing Filtering Strategy based on IP Prefix List

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} prefix-list < WORD > {in | out}

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} prefix-list < WORD > {in | out}

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

WORD: IP prefix list name.

【Description】

neighbor prefix-list: the command is used to set a routing filter policy based on the IP prefix list for peer/peer groups.

By default, no route filtering strategy based on IP prefix list.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 prefix-list test
```

18.3.52 BGP Update Message Removes Private AS

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} remove-private-AS

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} remove-private-AS

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor remote-private-AS: the command is used to configure sending BGP update messages without carrying a private autonomous system number.

By default, BGP update messages are sent with a private autonomous system number.

Private AS numbers range from 64512 to 65535.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 remove-private-AS
```

18.3.53 Route Mapping Strategy

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} route-map < WORD > {in | out}

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} route-map < WORD > {in | out}

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

WORD: IP prefix list name.

【Description】

neighbor route-map: the command is used to configure the route mapping policy to be applied when sending or receiving routes from a neighbor.

Default: no filtering policy based on routing map .

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 route-map test
```

18.3.54 Route Reflection

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} route-reflector-client

no neighbor {<A.B.C.D> | <X:X::X:X> | <WORD>} route-reflector-client

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor route-reflector-client: the command is used to configure native as the route reflector and peer/peer groups as the route reflector client.

By default: route reflector and its clients are not configured.

Route reflection is used to reduce the number of peers when the IBGP router inside the AS is larger, and its clients only exchange information with the route reflector, which handles the exchange of information between clients and with other IBGP and EBGP routers. This command configure itself as a route reflector and the specified peer (group) as its client. Note: this configuration can only be conducted within AS.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 route-reflector-client
Switch(config- router)#neighbor 3.3.3.3 remote-as 200
Switch(config- router)#neighbor 3.3.3.3 route-reflector-client
```

18.3.55 Router Server

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} route-server-client

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} route-server-client

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor route-server-client: the command is used to configure the native as the routing server and the peer/peer group as the routing server client.

By default: no routing server and its clients are configured.

Routing service is used to reduce the number of peers between AS routers in EBGP environment. When the clients exchange information through the routing server, the server transparently transmits the routing message to other clients.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 route-server-client
Switch(config- router)#neighbor 3.3.3.3 remote-as 300
Switch(config- router)#neighbor 3.3.3.3 route-server-client
```

18.3.56 Community Property Transfer

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} send-community [both | extended | standard]

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} send-community [both | extended | standard]

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

both: both standard and extended groups are transmitted

extended: only extended groups are transmitted.

standard: only standard groups are transmitted.

【Description】

neighbor send-community: the command is used to specify the transmission of group attributes to the specified BGP neighbor.

By default: group attributes is sent.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 send-community
```

18.3.57 Disable BGP Connection

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} shutdown

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} shutdown

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor shutdown: the command is used to close a BGP connection with a specified BGP peer.

By default: the connection is not closed.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 shutdown
```

18.3.58 Store BGP Original Routing Information

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} soft-reconfiguration inbound

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} soft-reconfiguration inbound

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor soft-reconfiguration: the command is used to allow the storage of original routing information from BGP peers.

By default: storage is not allowed.

When configured to allow the storage of original routing information from BGP peers, the system will first store the original routing information in the cache and apply it to the end immediately after restart, thus reducing the consumption of exchange with other routers. This command makes sense when the routing refresh capability is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router)#neighbor 1.1.1.1 remote-as 200
```

```
Switch(config-router)#neighbor 1.1.1.1 soft-reconfiguration
inbound
```

18.3.59 Strict Ability Matching Connection

【Command】

neighbor {<A.B.C.D> | <X:X::X:X> | <WORD>} strict-capability-match
no neighbor {<A.B.C.D> | <X:X::X:X> | <WORD>}
strict-capability-match

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor strict-capability-match: the command is used to configure a capability match that is strictly required when establishing a connection.

By default，strict capability matching is not configured.

This command is only valid for BGP multi-protocol. After the use of this command, the neighbor can only be established when the BGP multi-protocol capabilities of both parties match. Otherwise, it cannot be established. Whether or not other abilities are matched does not affect the establishment of neighbors.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router)#neighbor 1.1.1.1 remote-as 200
Switch(config-router)#neighbor 1.1.1.1 strict-capability-match
```

18.3.60 BGP Session Timer with the Specified Peer

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} timers { <keepalive> <holdtime> | connect <connect-time>}

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} timers { <keepalive> <holdtime> | connect <connect-time>}

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

keepalive: the time interval between sending KEEPALIVE messages to a specified BGP peer, range: 0-65535 seconds.

holdtime: specifies the interval between BGP peer sending KEEPALIVE messages, range: 0-65535 seconds.

connect-time: time interval for initiating a reconnection to a specified BGP peer, range: 1-65535 seconds.

【Description】

neighbor timers: the command is used to configure the alive time interval and hold times of BGP sessions established with a specified peer/peer group, as well as the reconnection interval.

By default, keepalive time is 60s, holdtime is 240s and connect-time is 120s.

Keepalives are recommended to be no more than one-third of the holdtime value.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router)#neighbor 1.1.1.1 remote-as 200
Switch(config-router)#neighbor 1.1.1.1 timers 50 150
```

18.3.61 Unsuppress the Routing Map

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} unsuppress-map <WORD>

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} unsuppress-map <WORD>

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

WORD: the name of the route-map.

【Description】

neighbor unsuppress-map: the command is used to configure or cancel unsuppress of cases that match the specified routing map.

By default, it is not configured.

This function is mainly for specific routes that are aggregated and summary-only suppressed. Routes that meet the routing mapping conditions are not suppressed in this case and are still issued separately.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 3 permit 100.1.3.0
Switch(config)#route-map test permit 10
Switch (config-route-map)#match ip address 3
Switch (config-route-map)#exit
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 unsuppress-map test
```

18.3.62 Specify the BGP Connection Interface

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} update-source < WORD >

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} update-source < WORD >

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

WORD: interface name or address.

【Description】

neighbor update-source: the command is used to set up the network interface to be used when establishing a BGP connection in a specified IBGP peer.

By default: use the best local interface as the output interface.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 update-source vlanif1
```

18.3.63 Configure the BGP Version of the Peer

【Command】

neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} version 4

no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} version

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

【Description】

neighbor version: the command is used to configure the BGP version number of the peer, currently only version 4 available.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#neighbor 1.1.1.1 remote-as 200
Switch(config- router)#neighbor 1.1.1.1 version 4
```

18.3.64 Configure the Weight Value of the Peer

【Command】

```
neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} weight <value>
no neighbor {<A.B.C.D > | <X:X::X:X> | <WORD>} weight <value>
```

【View】

BGP configuration view, IPv4 or IPv6 address family of BGP configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IPv4 peer interface address.

X:X::X:X: IPv6 peer interface address.

WORD: name of the peer group.

value: neighbor's weight. Value range is 0-65535.

【Description】

neighbor weight: the command is used to configure the weights of the peer.

By default: routes from other routers have a default weight of 0. The weight of the local static configuration defaults to 32768.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config-router)#neighbor 1.1.1.1 remote-as 200
Switch(config-router)#neighbor 1.1.1.1 weight 40
```

18.3.65 Introduce Additional Routing Information to the BGP

【Command】

redistribute { connected| static| ospf |rip} [route-map<word>]

no redistribute { connected| static| ospf| rip} [route-map<word>]

【View】

BGP View

【Default Level】

2: Configuration level

【Parameter】

connected：connected route

static: static route;

ospf：OSPF route.

rip：RIP route.

<word>：a pointer to the route map used to introduce routes.

【Description】

redistribute: the command introduces routes learned from other routing protocols into BGP.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#redistribute ospf
```

18.3.66 BGP Connection Survival Interval and Holdtime

【Command】

timers bgp <keepalive> <holdtime>

no timers basic

【View】

BGP View

【Default Level】

2: Configuration level

【Parameter】

keepalive: keepalive interval, the range is 0-65535 seconds.

holdtime: the range is 0-65535 seconds.

【Description】

timers bgp: the command is used to configure the intervals of the BGP connection alive time and hold time.

By default, the BGP connection has a live time interval of 60 seconds and a hold time of 180 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#router bgp 100
Switch(config- router)#timers bgp 80 200
```

18.3.67 Display BGP Route Information

【Command】

show ip bgp [<A.B.C.D >] [<A.B.C.D/M >]

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IP address.

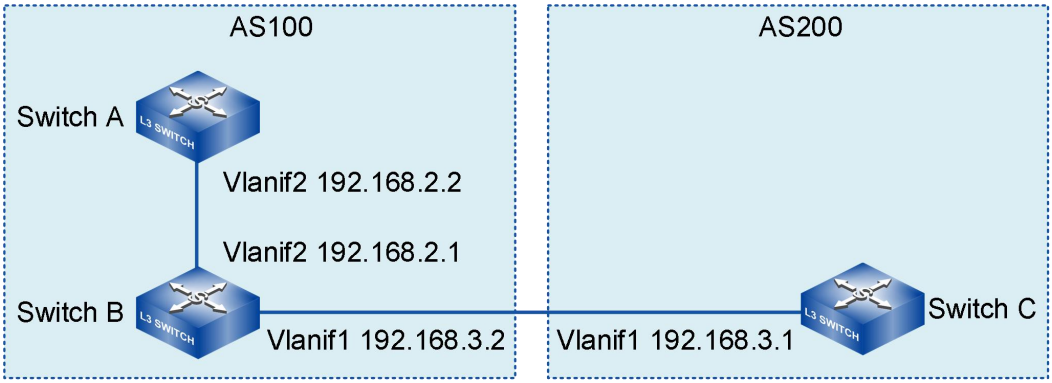
A.B.C.D/M: IP address and mask.

【Description】

show ip bgp: command is used to display routing information maintained by BGP.

【Instance】

Networking: As shown in the figure below, device A is connected to device B through VLANIF2, and device B is connected to device C through VLANIF1. Device C displays routing information maintained by BGP via command.




```
SwitchA> enable
SwitchA#configure terminal
SwitchA(config)#interface vlanif1
SwitchA(config- vlanif1)#ip address 192.168.1.1/24
SwitchA(config- vlanif1)#exit
SwitchA(config)#vlan database
SwitchA(config-vlan)#vlan 2
SwitchA(config-vlan)#exit
SwitchA(config)#ip interface vlan 2
SwitchA(config)#interface vlanif2
SwitchA(config- vlanif2)#ip address 192.168.2.2/24
SwitchA(config- vlanif2)#exit
SwitchA(config)#ip interface loopback 0
SwitchA(config)#interface loopback0
SwitchA(config- loopback0)#ip address 2.2.2.2/32
SwitchA(config- loopback0)#exit
SwitchA(config)#router ospf
SwitchA(config-router)#router-id 2.2.2.2
SwitchA(config-router)#network 192.168.1.0/24 area 0
SwitchA(config-router)#network 192.168.2.0/24 area 0
SwitchA(config-router)#network 2.2.2.2/32 area 0
SwitchA(config-router)#exit
SwitchA(config)#router bgp 100
SwitchA(config-router)#bgp router-id 2.2.2.2
SwitchA(config-router)#network 192.168.1.0/24
SwitchA(config-router)#network 192.168.2.0/24
SwitchA(config-router)#neighbor 1.1.1.1 remote-as 100

SwitchB> enable
SwitchB#configure terminal
SwitchB(config)#interface vlanif1
SwitchB(config- vlanif1)#ip address 192.168.3.2/24
SwitchB(config- vlanif1)#exit
SwitchB(config)#vlan database
```

```
SwitchB(config-vlan)#vlan 2
SwitchB(config-vlan)#exit
SwitchB(config)#ip interface vlan 2
SwitchB(config)#interface vlanif2
SwitchB(config- vlanif2)#ip address 192.168.2.1/24
SwitchB(config-vlanif2)#exit
SwitchB(config)#ip interface loopback 0
SwitchB(config)#interface loopback0
SwitchB(config-loopback0)#ip address 1.1.1.1/32
SwitchB(config-loopback0)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#router-id 1.1.1.1
SwitchB(config-router)#network 1.1.1.1/32 area 0
SwitchB(config-router)#network 192.168.2.0/24 area 0
SwitchB(config-router)#network 192.168.3.0/24 area 0
SwitchB(config-router)#exit
SwitchB(config)#router bgp 100
SwitchB(config-router)#bgp router-id 1.1.1.1
SwitchB(config-router)#network 192.168.1.0/24
SwitchB(config-router)#network 192.168.2.0/24
SwitchB(config-router)#neighbor 2.2.2.2 remote-as 100
SwitchB(config-router)#neighbor 3.3.3.3 remote-as 200

SwitchC> enable
SwitchC#configure terminal
SwitchC(config)#interface vlanif1
SwitchC(config- vlanif1)#ip address 192.168.3.1/24
SwitchC(config- vlanif1)#exit
SwitchC(config)#ip interface loopback 0
SwitchC(config)#interface loopback0
SwitchC(config- loopback0)#ip address 3.3.3.3/32
SwitchC(config- loopback0)#exit
SwitchC(config)#router ospf
SwitchC(config-router)#router-id 3.3.3.3
SwitchC(config-router)#network 192.168.3.0/24 area 0
SwitchC(config-router)#network 3.3.3.3/32 area 0
SwitchC(config- router)#exit
SwitchC(config)#router bgp 200
SwitchC(config-router)#bgp router-id 3.3.3.3
SwitchC(config-router)#network 192.168.3.0/24
SwitchC(config-router)#neighbor 1.1.1.1 remote-as 100
SwitchC(config-router)#end
SwitchC#show ip bgp
```

```
BGP table version is 1, local router ID is 3.3.3.3
Status codes: s suppressed, d damped, h history, * valid, > best,
i - internal,            S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric LocPrf Weight Path
*> 192.168.3.0      0.0.0.0                  100  32768 i

Total number of prefixes 1
```

18.3.68 Display BGP IPv4 Unicast Route Information

【Command】

show ip bgp ipv4 unicast [<A.B.C.D >] [<A.B.C.D/M >]

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IP address.

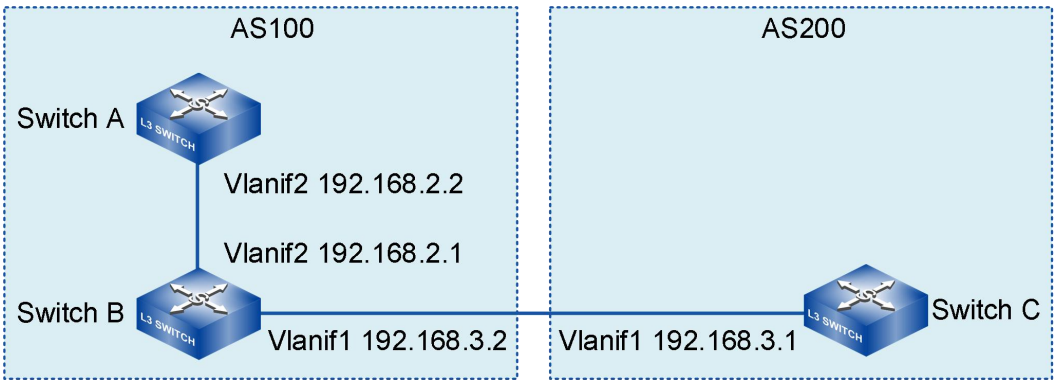
A.B.C.D/M: IP address and mask.

【Description】

show ip bgp: the command is used to display ipv4 unicast routing information maintained by BGP.

【Instance】

Networking: device A is connected to device B via VLANIF2, and device B is connected to device C via VLANIF1 as shown in the figure below. Device C displays the IPv4 unicast routing information maintained by BGP via command.



```
SwitchA> enable
SwitchA#configure terminal
SwitchA(config)#interface vlanif1
SwitchA(config- vlanif1)#ip address 192.168.1.1/24
SwitchA(config- vlanif1)#exit
SwitchA(config)#vlan database
SwitchA(config-vlan)#vlan 2
SwitchA(config-vlan)#exit
SwitchA(config)#ip interface vlan 2
SwitchA(config)#interface vlanif2
SwitchA(config- vlanif2)#ip address 192.168.2.2/24
SwitchA(config- vlanif2)#exit
SwitchA(config)#ip interface loopback 0
SwitchA(config)#interface loopback0
SwitchA(config- loopback0)#ip address 2.2.2.2/32
SwitchA(config- loopback0)#exit
SwitchA(config)#router ospf
SwitchA(config-router)#router-id 2.2.2.2
SwitchA(config-router)#network 192.168.1.0/24 area 0
SwitchA(config-router)#network 192.168.2.0/24 area 0
SwitchA(config-router)#network 2.2.2.2/32 area 0
SwitchA(config-router)#exit
SwitchA(config)#router bgp 100
SwitchA(config-router)#bgp router-id 2.2.2.2
SwitchA(config-router)#network 192.168.1.0/24
SwitchA(config-router)#network 192.168.2.0/24
SwitchA(config-router)#neighbor 1.1.1.1 remote-as 100

SwitchB> enable
SwitchB#configure terminal
SwitchB(config)#interface vlanif1
SwitchB(config- vlanif1)#ip address 192.168.3.2/24
SwitchB(config- vlanif1)#exit
SwitchB(config)#vlan database
SwitchB(config-vlan)#vlan 2
SwitchB(config-vlan)#exit
SwitchB(config)#ip interface vlan 2
SwitchB(config)#interface vlanif2
SwitchB(config- vlanif2)#ip address 192.168.2.1/24
SwitchB(config-vlanif2)#exit
SwitchB(config)#ip interface loopback 0
SwitchB(config)#interface loopback0
```

```
SwitchB(config- loopback0)#ip address 1.1.1.1/32
SwitchB(config- loopback0)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#router-id 1.1.1.1
SwitchB(config-router)#network 1.1.1.1/32 area 0
SwitchB(config-router)#network 192.168.2.0/24 area 0
SwitchB(config-router)#network 192.168.3.0/24 area 0
SwitchB(config-router)#exit
SwitchB(config)#router bgp 100
SwitchB(config-router)#bgp router-id 1.1.1.1
SwitchB(config-router)#network 192.168.1.0/24
SwitchB(config-router)#network 192.168.2.0/24
SwitchB(config-router)#neighbor 2.2.2.2 remote-as 100
SwitchB(config-router)#neighbor 3.3.3.3 remote-as 200

SwitchC> enable
SwitchC#configure terminal
SwitchC(config)#interface vlanif1
SwitchC(config- vlanif1)#ip address 192.168.3.1/24
SwitchC(config- vlanif1)#exit
SwitchC(config)#ip interface loopback 0
SwitchC(config)#interface loopback0
SwitchC(config- loopback0)#ip address 3.3.3.3/32
SwitchC(config- loopback0)#exit
SwitchC(config)#router ospf
SwitchC(config-router)#router-id 3.3.3.3
SwitchC(config-router)#network 192.168.3.0/24 area 0
SwitchC(config-router)#network 3.3.3.3/32 area 0
SwitchC(config- router)#exit
SwitchC(config)#router bgp 200
SwitchC(config-router)#bgp router-id 3.3.3.3
SwitchC(config-router)#network 192.168.3.0/24
SwitchC(config-router)#neighbor 1.1.1.1 remote-as 100
SwitchC(config-router)#end
SwitchC#show ip bgp ipv4 unicast
BGP table version is 1, local router ID is 3.3.3.3
Status codes: s suppressed, d damped, h history, * valid, > best,
i - internal,            S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network         Next Hop          Metric LocPrf Weight Path
*> 192.168.3.0     0.0.0.0                  100  32768 i
```

```
Total number of prefixes 1
```

18.3.69 Display BGP Summary Information

【Command】

show ip bgp summary

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IP address.

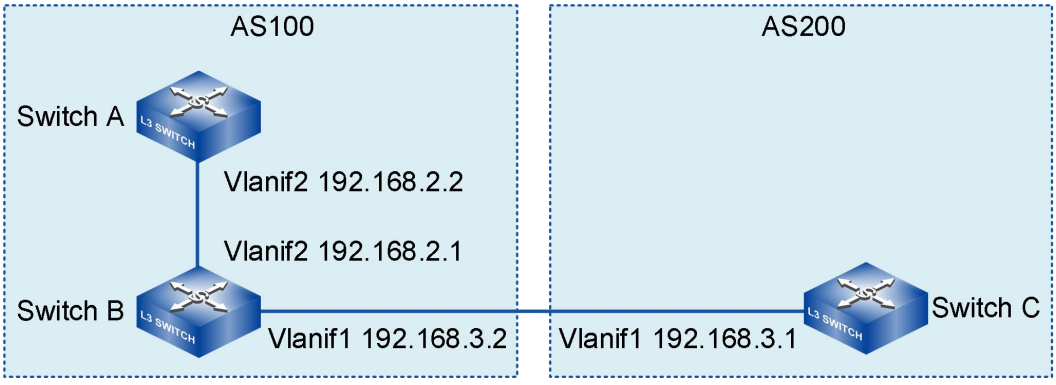
A.B.C.D/M: IP address and mask.

【Description】

show ip bgp summary: The command is used to display BGP summary information.

【Instance】

Networking: device A is connected to device B via VLANIF2, and device B is connected to device C via VLANIF1 as shown in the figure below. Device C displays BGP summary information via command.




```
SwitchA> enable
SwitchA#configure terminal
SwitchA(config)#interface vlanif1
SwitchA(config- vlanif1)#ip address 192.168.1.1/24
SwitchA(config- vlanif1)#exit
SwitchA(config)#vlan database
SwitchA(config-vlan)#vlan 2
SwitchA(config-vlan)#exit
SwitchA(config)#ip interface vlan 2
```

```
SwitchA(config)#interface vlanif2
SwitchA(config- vlanif2)#ip address 192.168.2.2/24
SwitchA(config- vlanif2)#exit
SwitchA(config)#ip interface loopback 0
SwitchA(config)#interface loopback0
SwitchA(config- loopback0)#ip address 2.2.2.2/32
SwitchA(config- loopback0)#exit
SwitchA(config)#router ospf
SwitchA(config-router)#router-id 2.2.2.2
SwitchA(config-router)#network 192.168.1.0/24 area 0
SwitchA(config-router)#network 192.168.2.0/24 area 0
SwitchA(config-router)#network 2.2.2.2/32 area 0
SwitchA(config-router)#exit
SwitchA(config)#router bgp 100
SwitchA(config-router)#bgp router-id 2.2.2.2
SwitchA(config-router)#network 192.168.1.0/24
SwitchA(config-router)#network 192.168.2.0/24
SwitchA(config-router)#neighbor 1.1.1.1 remote-as 100

SwitchB> enable
SwitchB#configure terminal
SwitchB(config)#interface vlanif1
SwitchB(config- vlanif1)#ip address 192.168.3.2/24
SwitchB(config- vlanif1)#exit
SwitchB(config)#vlan database
SwitchB(config-vlan)#vlan 2
SwitchB(config-vlan)#exit
SwitchB(config)#ip interface vlan 2
SwitchB(config)#interface vlanif2
SwitchB(config- vlanif2)#ip address 192.168.2.1/24
SwitchB(config-vlanif2)#exit
SwitchB(config)#ip interface loopback 0
SwitchB(config)#interface loopback0
SwitchB(config- loopback0)#ip address 1.1.1.1/32
SwitchB(config- loopback0)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#router-id 1.1.1.1
SwitchB(config-router)#network 1.1.1.1/32 area 0
SwitchB(config-router)#network 192.168.2.0/24 area 0
SwitchB(config-router)#network 192.168.3.0/24 area 0
SwitchB(config-router)#exit
SwitchB(config)#router bgp 100
SwitchB(config-router)#bgp router-id 1.1.1.1
```

```
SwitchB(config-router)#network 192.168.1.0/24
SwitchB(config-router)#network 192.168.2.0/24
SwitchB(config-router)#neighbor 2.2.2.2 remote-as 100
SwitchB(config-router)#neighbor 3.3.3.3 remote-as 200

SwitchC> enable
SwitchC#configure terminal
SwitchC(config)#interface vlanif1
SwitchC(config- vlanif1)#ip address 192.168.3.1/24
SwitchC(config- vlanif1)#exit
SwitchC(config)#ip interface loopback 0
SwitchC(config)#interface loopback0
SwitchC(config- loopback0)#ip address 3.3.3.3/32
SwitchC(config- loopback0)#exit
SwitchC(config)#router ospf
SwitchC(config-router)#router-id 3.3.3.3
SwitchC(config-router)#network 192.168.3.0/24 area 0
SwitchC(config-router)#network 3.3.3.3/32 area 0
SwitchC(config- router)#exit
SwitchC(config)#router bgp 200
SwitchC(config-router)#bgp router-id 3.3.3.3
SwitchC(config-router)#network 192.168.3.0/24
SwitchC(config-router)#neighbor 1.1.1.1 remote-as 100
SwitchC(config-router)#end
SwitchC#show ip bgp summary
BGP router identifier 3.3.3.3, local AS number 200
BGP table version is 1
1 BGP AS-PATH entries
0 BGP community entries

Neighbor    V     AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down
State/PfxRcd
1.1.1.1     4   100       0      53        0    0    0    never
Active

Total number of neighbors 1
```

19 IPv6 Configuration

19.1 Overview

Internet Protocol version 6 (IPv6), also called IP Next Generation (IPng), is the second-generation network layer protocol. Designed by the Internet Engineering Task Force (IETF), IPv6 is an upgraded version of Internet Protocol version 4 (IPv4).

IPv6 was developed in response to rapidly increasing Internet use. IPv4, being easy to implement, simple to use, and providing good interoperability, had developed quickly at the early stage of internet development. However, with the rapid development of the Internet, the shortcomings of IPv4 design have become increasingly obvious. The emergence of IPv6 has solved some disadvantages of IPv4. This is mainly due to IPv4 address exhaustion.

- Address space

  Deficiency in IPv4:

  - IPv4 addresses are 32 bits long, theoretically giving an available IP address space that contains about 4.3 billion IP addresses. The currently available IP addresses are no longer sufficient to continually support the rapid expansion of the Internet. IPv4 address resources are allocated unevenly. USA address resources account for almost half of the global address space, with barely enough addresses left for Europe, and still fewer for the Asia-Pacific area. Furthermore, the development of mobile IP and broadband technologies still requires more IP addresses. The process of IP addresses being used up is known as IP address exhaustion.
  - There have been several solutions to the IPv4 address shortage. The most representative ones are Classless Inter-domain Routing (CIDR) and Network Address Translator (NAT). These disadvantages prompted the development of IPv6.

Advantage of IPv6:

- IPv6 addresses are 128 bits long. A 128 bit structure allows for an address space of 2128 (4.3 billion x 4.3 billion x 4.3 billion x 4.3 billion) possible addresses. This vast address space makes it very unlikely that IPv6 address exhaustion will ever occur.

- Packet format

  Deficiency in IPv4:

  - The IPv4 packet header carries the Options field, including security, timestamp, and record route options. The variable length of the Options field makes the IPv4 packet header length range from 20 bytes to 60 bytes. IPv4 packets often need to be forwarded by intermediate devices. Therefore, using the Options field occupies a large amount of resources, which means this field is rarely used in practice.

  Advantage of IPv6:

  - Unlike the IPv4 packet header, the IPv6 packet header does not carry IHL, identifier, flag, fragment offset, header checksum, option, or padding fields, but it does carry the flow label field. This facilitates IPv6 packet processing and improves processing efficiency. To support various options without changing the existing packet format, the Extension Header information field is added to the IPv6 packet header, improving IPv6 flexibility.

- Autoconfiguration and readdressing

  Deficiency in IPv4:

  - IP addresses often need to be reallocated during network expansion or re-planning. Currently, IPv4 depends on Dynamic Host Configuration Protocol (DHCP) to provide address autoconfiguration and readdressing to simplify address maintenance.

  Advantage of IPv6:

  - IPv6 provides address autoconfiguration to allow hosts to automatically discover networks and obtain IPv6 addresses, improving network manageability.

- Route summarization

  Deficiency in IPv4:

  - Many non-contiguous IPv4 addresses are allocated. Routes cannot be summarized effectively due to incorrect IPv4 address allocation and planning. The increasingly large routing table consumes a lot of memory and affects forwarding efficiency. Manufacturers must continually upgrade devices to improve route addressing and forwarding performance.

Advantage of IPv6:

- The enormous address space allows for the hierarchical network design in IPv6. The hierarchical network design in IPv6 facilitates route summarization and improves forwarding efficiency.

- End-to-end security support

  Deficiency in IPv4:

  - The original IPv4 framework does not support end-to-end security because security was not fully considered during the initial design.

  Advantage of IPv6:

  - IPv6 supports IP Security (IPSec) authentication and encryption at the network layer, providing end-to-end security.

- Quality of Service (QoS) support

  Deficiency in IPv4:

  - When regarding network conferencing, network telephones, and network TVs, customer demands better QoS to support the real-time forwarding of voice, data, and video services. IPv4 doesn’t have special means to support QoS.

  Advantage of IPv6:

  - The Flow Label field in IPv6 guarantees QoS for voice, data, and video services.

- Support for mobility

  Deficiency in IPv4:

  - Due to the development of the Internet, mobile IPv4 experiences significant problems such as triangular routing and source address filtering.

  Advantage of IPv6:

  - Mobile IPv6 improves mobile communication efficiency and is transparent to the application layer because IPv6 has the native capability to support mobility. Unlike mobile IPv4, mobile IPv6 uses the neighbor discovery function to discover a foreign network and obtain a care-of address without using any foreign agent. The mobile node and peer node can communicate using the routing header and destination options header. This function solves the problems of triangular routing and source address filtering found in mobile IPv4.

19.2 Principle Description

19.2.1 IPV6 Address

19.2.1.1 IPv6 Address Representation

An IPv6 address is 128 bits long and is written as eight groups of four hexadecimal digits (base 16 digits represented by the numbers 0-9 and the letters A-F). Each group is separated by a colon (:). For example, FC00:0000:130F:0000:0000:09C0:876A:130B is a complete and valid IPv6 address. For convenience, IPv6 addresses can be written in a compressed format. Taking the IPv6 address FC00:0000:130F:0000:0000:09C0:876A:130B as an example:

- Any leading zeroes in a group can be omitted. The example address now becomes FC00:0:130F:0:0:9C0:876A:130B.
- A double colon (::) can be used when two or more consecutive groups contain all zeros. The example address now becomes FC00:0:130F::9C0:876A:130B.

Note

An IPv6 address can contain only one double colon (::). Otherwise, a computer cannot determine the number of zeros in a group when restoring the compressed address to the original 128-bit address.

19.2.1.2 IPv6 Address Structure

IPv6 addresses have two parts:

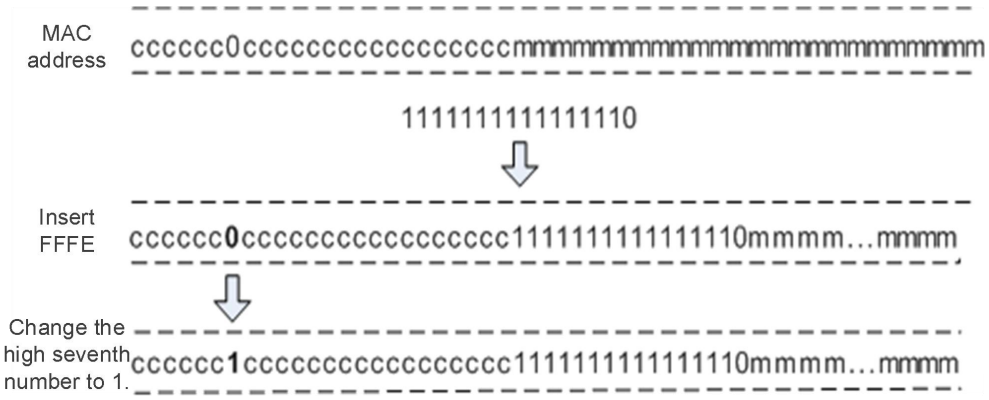
- Network prefix: Corresponds to the network ID of an IPv4 address. It is comprised of n bits；
- Interface identifier (interface ID): Corresponds to the host ID of an IPv4 address. It is comprised of 128-n bits.

Note

If the first 3 bits of an IPv6 unicast address are not 000, the interface ID must contain 64 bits. If the first 3 bits are 000, there is no such limitation.

You can manually configure the interface ID, generate it through system software, or generate it in IEEE 64-bit Extended Unique Identifier (EUI-64) format. Generating an interface ID in EUI-64 format is the most common practice.

IEEE EUI-64 standards convert an interface MAC address into an IPv6 interface ID. The Figure below shows a 48-bit MAC address. When used as an interface ID, the first 24 bits (expressed by c) are a vendor identifier, and the last 24 bits (expressed by m) are an extension identifier. If the higher seventh bit is 0, the MAC address is locally unique. During conversion, EUI-64 inserts FFFE between the vendor identifier and extension identifier. The higher seventh bit also changes from 0 to 1 to indicate that the interface ID is globally unique.



For example, if the MAC address is 000E-0C82-C4D4, the interface ID is 020E:0CFF:FE82:C4D4 after the conversion.

Converting MAC addresses into IPv6 interface IDs reduces the configuration workload. When using stateless address autoconfiguration, you only need an IPv6 network prefix to obtain an IPv6 address. One defect of this method, however, is that an IPv6 address is easily calculable based on a MAC address, and could therefore be used for malicious attacks.

19.2.1.3 IPv6 Address Classification

IPv6 addresses can be classified as unicast, multicast, or a new class called anycast. Unlike IPv4, there is no broadcast IPv6 address. Instead, a multicast address can be used as a broadcast address.

IPv6 Unicast Address

An IPv6 unicast address identifies an interface. Since each interface belongs to a node, the IPv6 unicast address of any interface can identify the relevant node.

Packets sent to an IPv6 unicast address are delivered to the interface identified by that address.

IPv6 defines multiple types of unicast addresses, including the unspecified address, loopback address, global unicast address, link-local address, and unique local address.

- unspecified address

  The IPv6 unspecified address is 0:0:0:0:0:0:0:0/128 or ::/128, indicating that an interface or a node does not have an IP address. It can be used as the source IP address of some packets, such as Neighbor Solicitation (NS) messages, in duplicate address detection. Devices do not forward packets with an unspecified address as the source IP address.

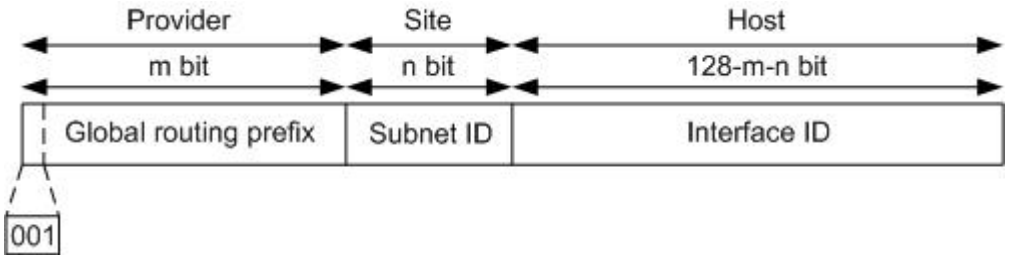
- Loopback address

  The IPv6 loopback address is 0:0:0:0:0:0:0:1/128 or ::1/128. Similar to the IPv4 loopback address 127.0.0.1, the IPv6 loopback address is used when a node needs to send IPv6 packets to itself. This IPv6 loopback address is usually used as the IP address of a virtual interface, such as a loopback interface. The loopback address cannot be used as the source or destination IP address of packets needing to be forwarded.

- Group unicast address

  An IPv6 global unicast address is an IPv6 address with a global unicast prefix, which is similar to an IPv4 public address. IPv6 global unicast addresses support route prefix summarization, helping limit the number of global routing entries.

  The Figure below shows a global unicast address consisting of a global routing prefix, subnet ID, and interface ID.



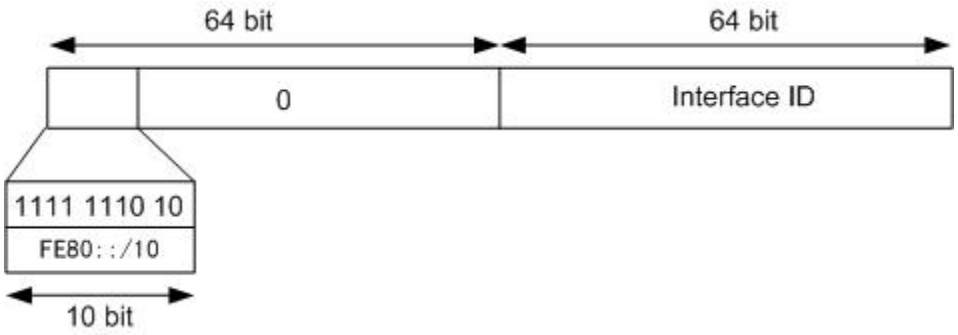
  - Global routing prefix: is assigned by a service provider to an organization. A global routing prefix is comprised of at least 48 bits. Currently, the first 3 bits of every assigned global routing prefix is 001.
  - Subnet ID Subnet ID is used by organizations to construct a local network (site). There are a maximum of 64 bits for subnet ID. Subnet ID has similar function as an IPv4 subnet number.
  - Interface ID: identifies an interface. Used to identify a device (host).

- Link local address

  Link-local addresses are used only in communication between nodes on the same local link. A link-local address uses a link-local prefix of FE80::/10 as the first 10 bits (1111111010 in binary) and an interface ID as the last 64 bits.

  When IPv6 runs on a node, a link-local address that consists of a fixed prefix and an interface ID in EUI-64 format is automatically assigned to each interface of the node. This mechanism enables two IPv6 nodes on the same link to communicate without any configuration. It makes link-local addresses widely used in neighbor discovery and stateless address configuration.

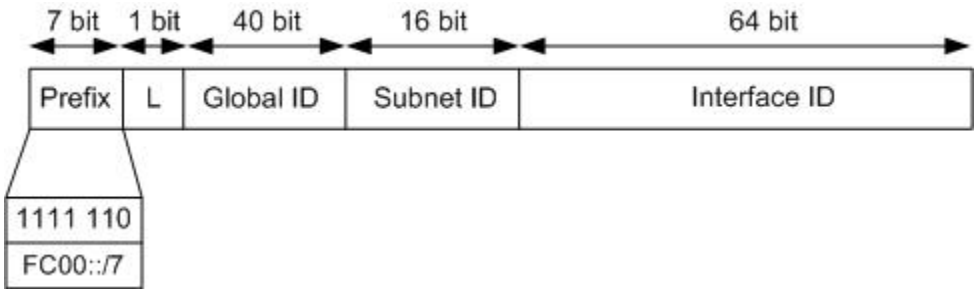
  Devices do not forward IPv6 packets with the link-local address as a source or destination address to devices on different links. The link-local address format is shown in the following figure:



- Unique local address

  Unique local addresses are used only within a site. Site-local addresses, according to RFC 3879, have been replaced by unique local addresses.

  Unique local addresses are similar to IPv4 private addresses. Any organization that does not obtain a global unicast address from a service provider can use a unique local address. However, the unique local addresses are routable only within a local network, not the Internet as a whole. The unique local address format is shown in the following figure:



  - Prefix: is fixed as FC00::/7.
  - L: is set to 1 if the address is valid within a local network. The value 0 is reserved for future expansion.

- Global ID: indicates a globally unique prefix, which is pseudo-randomly allocated (for details, see RFC 4193).
- Subnet ID: identifies a subnet within the site.
- Interface ID: identifies an interface.

A unique local address has the following features:

- Has a globally unique prefix that is pseudo-randomly allocated with a high probability of uniqueness.
- Allows private connections between sites without creating address conflicts.
- Has a well-known prefix (FC00::/7) that allows for easy route filtering at site boundaries.
- Does not conflict with any other addresses if it is accidentally routed offsite.
- Functions as a global unicast address to applications.
- Is independent of Internet Service Providers (ISPs).

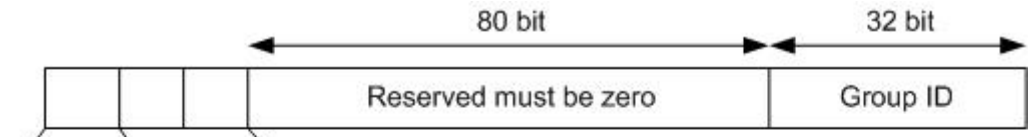
IPv6 Multicast Address

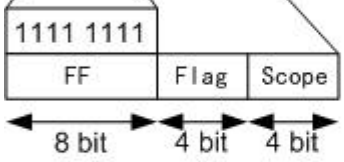
Like IPv4 multicast addresses, IPv6 multicast addresses identify groups of interfaces, which usually belong to different nodes. A node may belong to any number of multicast groups. Packets sent to an IPv6 multicast address are delivered to all the interfaces identified by the multicast address. For example, the multicast address FF02::1 indicates all nodes within the link-local scope, and FF02::2 indicates all routers within the link-local scope.

An IPv6 multicast address is composed of a prefix, a flag, a scope, and a group ID (global ID).

- Prefix: is fixed as FF00::/8.
- Flag: is 4 bits long. The high-order 3 bits are reserved and must be set to 0s. The last bit 0 indicates a permanently-assigned, well-known multicast address allocated by the Internet Assigned Numbers Authority (IANA). The last bit 1 indicates a non-permanently-assigned (transient) multicast address.
- Scope: is 4 bits long. It limits the scope where multicast data flows are sent on the network. Figure shows the field values and meanings.
- Group ID (global ID): is 112 bits long. It identifies a multicast group. RFC 2373 does not define all the 112 bits as a group ID but recommends using the low-order 32 bits as the group ID and setting all of the remaining 80 bits to 0s. In this case, each multicast group ID maps to a unique Ethernet multicast MAC address (for details, see RFC 2464).

Figure shows the IPv6 multicast address format.





field	value	description
Flag	1	temporary multicast address
	0	permanent multicast address
Scope	1	node
	2	link
	4	management
	5	site
	8	organization
	E	global
	the rest	unsigned or reserved

- Solicited-node Multicast Address

  A solicited-node multicast address is generated using an IPv6 unicast or anycast address of a node. When a node has an IPv6 unicast or anycast address, a solicited-node multicast address is generated for the node, and the node joins the multicast group that corresponds to its IPv6 unicast or anycast address. Each unicast or anycast address corresponds to a single solicited-node multicast address. It is often used in neighbor discovery and duplicate address detection. IPv6 does not support broadcast addresses or Address Resolution Protocol (ARP). In IPv6, Neighbor Solicitation (NS) packets are used to resolve IP addresses to MAC addresses. In IPv6, this function is accomplished by neighbor solicitation (NS) message. When a node needs to resolve the MAC address corresponding to an IPv6 address, it will send an NS message, and the destination IP of the message is the multicast address of the requested node corresponding to the IPv6 address to be resolved. Only the node with the multicast address will check the processing.

  The solicited-node multicast address consists of the prefix FF02::1:FF00:0/104 and the last 24 bits of the corresponding unicast address.

IPv6 Anycast Address

An anycast address identifies a group of network interfaces, which usually belong to different nodes. Packets sent to an anycast address are delivered to the nearest interface that is identified by the anycast address, depending on the routing protocols. Anycast addresses implement redundancy backup and load balancing functions when multiple hosts or nodes are provided with the same services. Currently, a unicast address is assigned to more than one interface to make a unicast address become an anycast address. When sending data packets to anycast addresses, senders cannot

determine which of the assigned devices will receive the packets. Which device receives the packets depends on the routing protocols running on the network. Anycast addresses are used in stateless applications, such as Domain Name Service (DNS).

IPv6 anycast addresses are allocated from the unicast address space. Mobile IPv6 applications also use anycast addresses.

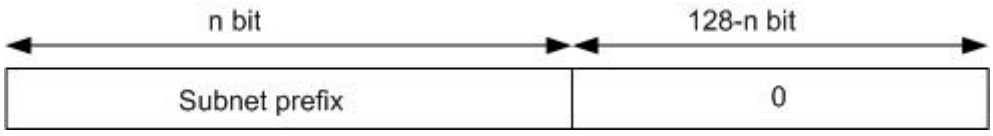
Note

IPv6 anycast addresses can be assigned only to routing devices but not hosts. Anycast addresses cannot be used as the source IP addresses of IPv6 packets.

- Subnet-router Anycast Address

  RFC 3513 predefines a subnet-router anycast address. Packets sent to a subnet-router anycast address are delivered to the nearest device on the subnet identified by the anycast address, depending on the routing protocols. All devices must support subnet-router anycast addresses. A subnet-router anycast address is used when a node needs to communicate with any of the devices on the subnet identified by the anycast address. For example, a mobile node needs to communicate with one of the mobile agents on the home subnet.

  In a subnet-router anycast address, the n-bit subnet prefix identifies a subnet, and the remaining bits are padded with 0s. The format is shown in the following figure：



19.2.2 IPv6 Packet Format

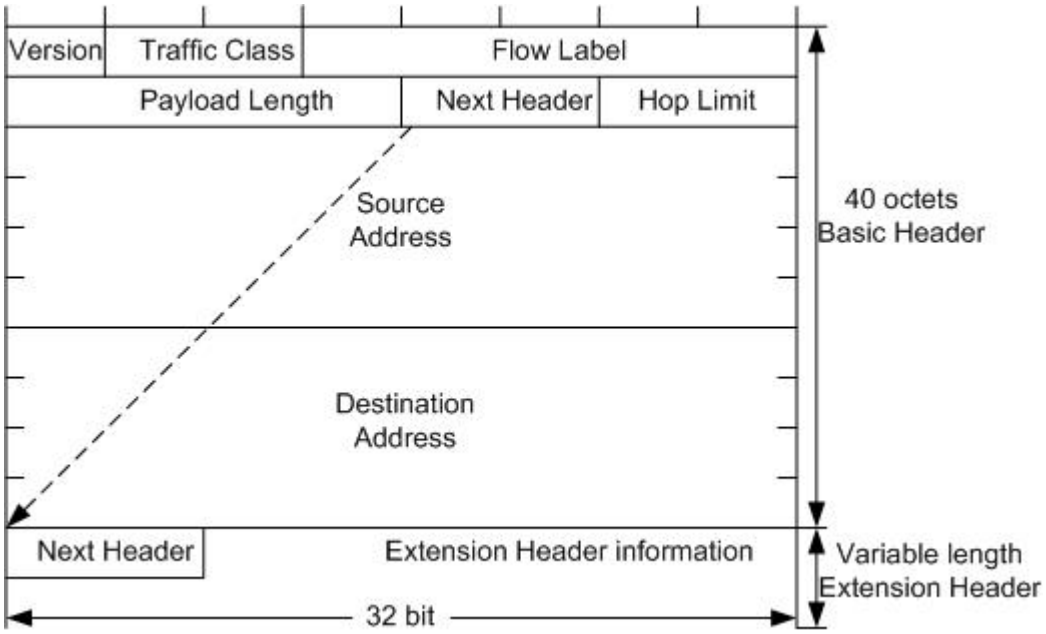
An IPv6 packet has three parts: an IPv6 basic header, one or more IPv6 extension headers, and an upper-layer protocol data unit (PDU).

An upper-layer PDU is composed of the upper-layer protocol header and its payload, which maybe an ICMPv6 packet, a TCP packet, or a UDP packet.

19.2.2.1 IPv6 Basic Header

An IPv6 basic header is fixed as 40 bytes long and has eight fields. Each IPv6 packet must have an IPv6 basic header that provides basic packet forwarding information, and which all devices parse on the forwarding path.

Figure shows the IPv6 basic header.



An IPv6 basic header contains the following fields:

- Version: 4 bits long. In IPv6, the value of the Version field is set to 6.
- Traffic Class: 8 bits long. This field indicates the class or priority of an IPv6 packet. The Traffic Class field is similar to the TOS field in an IPv4 packet and is mainly used in QoS control.
- Flow Label: 20 bits long. This field was added in IPv6 to differentiate traffic. A flow label and source IP address identify a data flow. Intermediate network devices can effectively differentiate data flows based on this field.
- Payload Length: 16 bits long. This field indicates the length of the IPv6 payload in bytes. The payload is the part of the IPv6 packet following the IPv6 basic header, including the extension header and upper-layer PDU. This field has a maximum value of 65535. If the payload length exceeds 65535 bytes, the field is set to 0, and the Jumbo Payload option in the Hop-by-Hop Options header is used to express the actual payload length.
- Next Header: 8 bits long. This field identifies the type of the first extension header that follows the IPv6 basic header or the protocol type in the upper-layer PDU.
- Hop Limit: 8 bits long. This field is similar to the Time to Live field in an IPv4 packet, defining the maximum number of hops that an IP packet can pass through. Each device that forwards the packet decrements the field value by 1. If

the field value is reduced to 0, the packet is discarded.

- Source Address: 128 bits long. This field indicates the address of the packet originator.
- Destination Address: 128 bits long. This field indicates the address of the packet recipient.

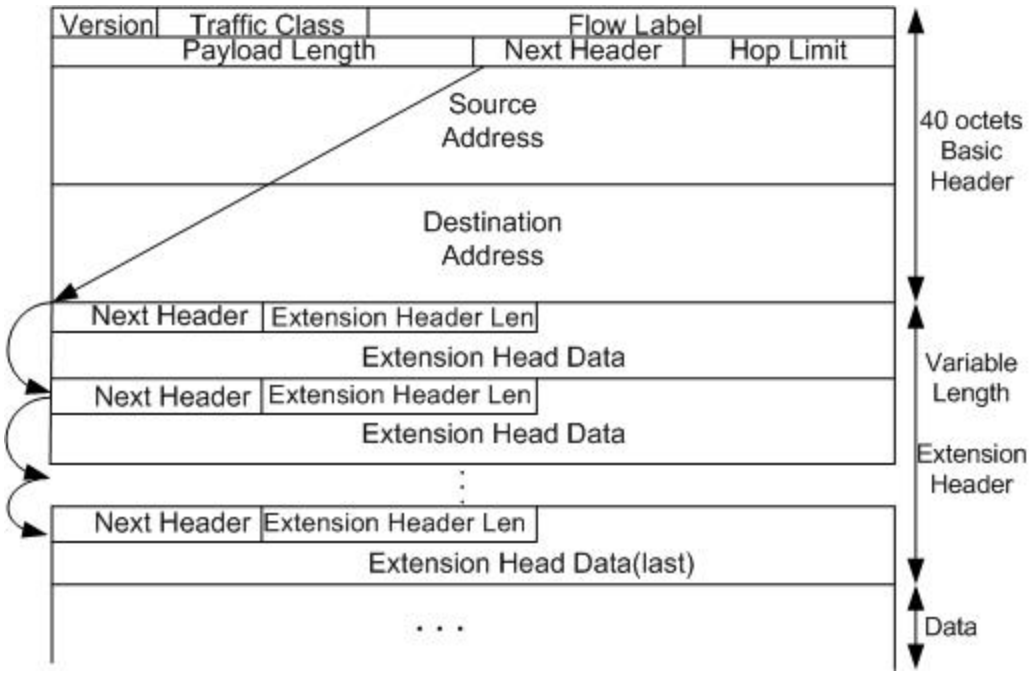
Unlike the IPv4 packet header, the IPv6 packet header does not carry IHL, identifier, flag, fragment offset, header checksum, option, or padding fields, but it does carry the flow label field. This facilitates IPv6 packet processing and improves processing efficiency. To support various options without changing the existing packet format, the Extension Header information field is added to the IPv6 packet header, improving flexibility. The following paragraphs describe IPv6 extension headers.

19.2.2.2 IPv6 Extension Header

An IPv4 packet header has an optional field (Options), which includes security, timestamp, and record route options. The variable length of the Options field makes the IPv4 packet header length range from 20 bytes to 60 bytes. When devices forward IPv4 packets with the Options field, many resources need to be used. Therefore, these IPv4 packets are rarely used in practice.

To improve packet processing efficiency, IPv6 uses extension headers to replace the Options field in the IPv4 header. Extension headers are placed between the IPv6 basic header and upper-layer PDU. An IPv6 packet may carry zero or more extension headers. The sender of a packet adds one or more extension headers to the packet only when the sender requests the destination device or other devices to perform special handling. Unlike IPv4, IPv6 has variable-length extension headers, which are not limited to 40 bytes. This facilitates further extension. To improve extension header processing efficiency and transport protocol performance, IPv6 requires that the extension header length be an integer multiple of 8 bytes.

When multiple extension headers are used, the Next Header field of an extension header indicates the type of the next header following this extension header. The Next Header field in the IPv6 basic header indicates the type of the first extension header, and the Next Header field in the first extension header indicates the type of the next extension header. If there are no extension headers following the current one, the Next Header field indicates the upper-layer protocol type. Figure shows the extension header format.



An IPv6 extension header contains the following fields:

- Next Header: 8 bits long. This is similar to the Next Header field in the IPv6 basic header. It indicates the type of the next extension header (if any) or the upper-layer protocol type.
- Extension Header Len: 8 bits long. This indicates the extension header length excluding the Next Header field.
- Extension Head Data: Variable length. This includes a series of options and the padding field.

RFC 2460 defines six IPv6 extension headers: Hop-by-Hop Options header, Destination Options header, Routing header, Fragment header, Authentication header, and Encapsulating Security Payload header.

Header Type	Next Header Field Value	Description
Hop-by-Hop Options header	0	This header carries information that every node must examine along the delivery path of a packet. This header is used in the following applications: • Jumbo payload (if the payload length exceeds 65535 bytes) • Prompting devices to check this option before the devices forward packets. • Resource Reservation Protocol (RSVP)

Header Type	Next Header Field Value	Description
Destination Options header	60	This header carries information that only the destination node of a packet examines. Currently, this header is used in mobile IPv6.
Routing header	43	An IPv6 source node uses this header to specify the intermediate nodes that a packet must pass through on the way to its destination. This option is similar to the Loose Source and Record Route option in IPv4.
Fragment header	44	Like IPv4 packets, the length of IPv6 packets to be forwarded cannot exceed the maximum transmission unit (MTU). When the packet length exceeds the MTU, the packet needs to be fragmented. In IPv6, the Fragment header is used by an IPv6 source node to send a packet larger than the MTU.
Authentication header	51	IPSec uses this header to provide data origin authentication, data integrity check, and packet anti-replay functions. It also protects some fields in the IPv6 basic header.
Encapsulating Security Payload header	50	This header provides the same functions as the Authentication header plus IPv6 packet encryption.

Conventions for IPv6 extension headers

When a single packet uses more than one extension header, the headers must be listed in the following order:

- IPv6 Basic Header

- Hop-by-Hop Options header
- Destination Options header
- Routing header
- Fragment header
- Authentication header
- Encapsulating Security Payload header
- Destination Options header
- Upper-layer header

Intermediate devices determine whether to process extension headers based on the Next Header field value in the IPv6 basic header. The intermediate devices do not need to examine or process all extension headers.

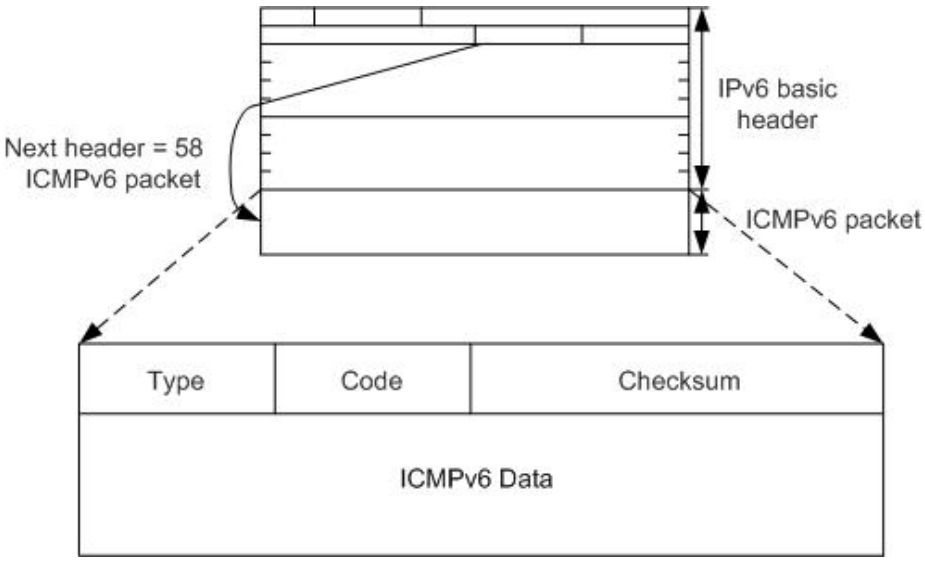
Each extension header can only occur once in an IPv6 packet, except for the Destination Options header which may occur twice (once before a Routing header and once before the upper-layer header).

19.2.3 ICMPv6

The Internet Control Message Protocol version 6 (ICMPv6) is one of the basic IPv6 protocols.

In IPv4, ICMP reports IP packet forwarding information and errors to the source node. ICMP defines certain messages such as Destination Unreachable, Packet Too Big, Time Exceeded, Echo Request, and Echo Reply to facilitate fault diagnosis and information management. ICMPv6 provides additional mechanisms alongside the current ICMPv4 functions such as Neighbor Discovery (ID), stateless address configuration (including duplicate address detection), and Path Maximum Transmission Unit (PMTU) discovery.

The protocol number of ICMPv6 (that is, the value of the Next Header field in an IPv6 packet) is 58. The message format of ICMPv6 is shown in the figure below：



Some fields in the packet are described as follows:

- Type: specifies a message type. Values 0 to 127 indicate the error message type, and values 128 to 255 indicate the informational message type.
- Code: indicates a specific message type.
- Checksum: indicates the checksum of an ICMPv6 packet.

19.2.3.1 Classification of ICMPv6 Error Messages

ICMPv6 error messages are generated when errors occur during IPv6 packet forwarding. They are classified into the following four types:

- Destination Unreachable message

  During IPv6 packet forwarding, if an IPv6 node detects that the destination address of a packet is unreachable, it sends an ICMPv6 Destination Unreachable message to the source node carrying information about the causes for the error message.

  In an ICMPv6 Destination Unreachable message, the value of the Type field is 1. Depending on the cause, the value of the Code field can be:

  - 0: No route to the destination device.
  - 1: Communication with the destination device is administratively prohibited.
  - 2: Not assigned.
  - 3: Destination IP address is unreachable.
  - 4: Destination port is unreachable.
- Packet Too Big message

  If an IPv6 node, during IPv6 packet forwarding, detects that the size of a packet exceeds the link MTU of the outbound interface, it sends an ICMPv6 Packet Too Big message to the source node. The link MTU of the outbound interface is

carried in the message. PMTU discovery is implemented based on Packet Too Big messages.

In a Packet Too Big message, the Type field value is 2 and the Code field value is 0.

- Time Exceeded message

  During the transmission of IPv6 packets, when a device receives a packet with a hop limit of 0 or a device reduces the hop limit to 0, it sends an ICMPv6 Time Exceeded message to the source node. During the processing of a packet to be fragmented and reassembled, an ICMPv6 Time Exceeded message is also generated when the reassembly time is longer than the specified period.

  In a Time Exceeded message, the Type field value is 3. Depending on the cause, the Code field value can be:

  - 0: Hop limit exceeded in packet transmission.
  - 1: Fragment reassembly timeout.

- Parameter Problem message

  When a destination node receives an IPv6 packet, it checks the validity of the packet. If it detects an error, it sends an ICMPv6 Parameter Problem message to the source node.

  In a Parameter Problem message, the Type field value is 4. Depending on the cause, the Code field value can be:

  - 0: A field in the IPv6 basic header or extension header is incorrect.
  - 1: The Next Header field in the IPv6 basic header or extension header cannot be identified.
  - 2: Unknown options exist in the extension header.

19.2.3.2 Classification of ICMPv6 Information Messages

ICMPv6 information messages provide diagnosis and additional host functions such as Multicast Listener Discovery (MLD) and Neighbor Discovery (ND). Common ICMPv6 information messages include Ping messages, which consist of Echo Request and Echo Reply messages.

- Echo Request messages: Echo Request messages are sent to destination nodes. After receiving an Echo Request message, the destination node responds with an Echo Reply message. In an Echo Request message, the Type field value is 128 and the Code field value is 0.
- Echo Reply messages: After receiving an Echo Request message, the destination node responds with an Echo Reply message. In an Echo Reply message, the Type field value is 129 and the Code field value is 0.

19.2.4 Neighbor Discovery

The Neighbor Discovery Protocol (NDP) is an enhancement of Address Resolution Protocol (ARP) and Internet Control Management Protocol (ICMP) router discovery in IPv4. In addition to ICMPv6 address resolution, NDP also provides the neighbor unreachable detection, duplicate address detection, router discovery, and redirection functions.

19.2.4.1 Address resolution

In IPv4, a host needs to obtain the link-layer address of the destination host through the ARP protocol for communication. In IPv6, the function of resolving from IP address to link layer address is also needed. Neighbor discovery protocol realizes this function.

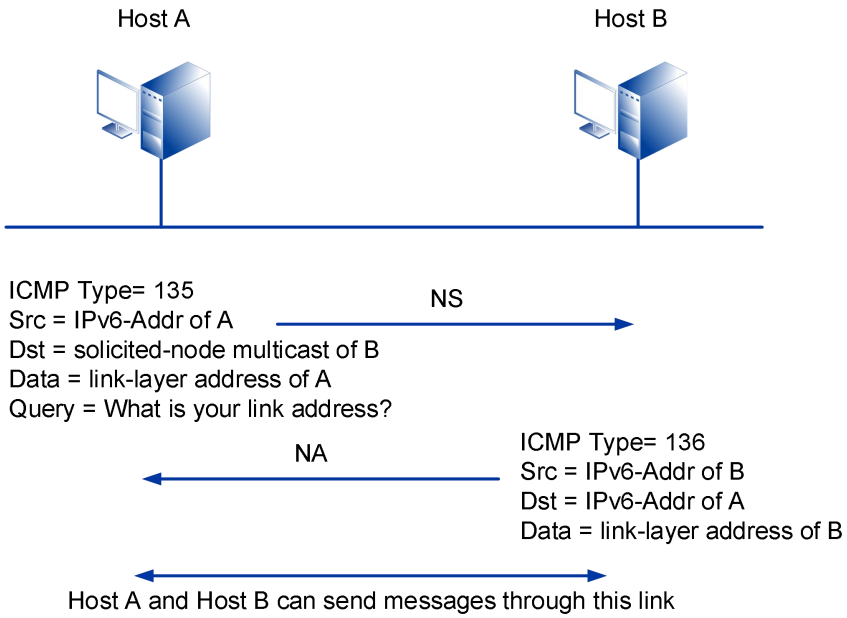
ARP messages are directly encapsulated in Ethernet messages, and the Ethernet protocol type is 0x0806. It is generally believed that ARP is positioned as a layer 2.5 protocol. The ND itself is implemented based on ICMPv6, ND the Ethernet protocol type is 0x86DD, that is, IPv6 message, ND the next header field value of IPv6 is 58, which indicates ICMPv6 message. since all messages used by nd protocol are encapsulated in ICMPv6 message, nd is generally regarded as layer 3 protocol. Completing address resolution at the third layer mainly brings the following benefits:

- Layer 3 address resolution enables Layer 2 devices to use the same address resolution protocol.
- Layer 3 security mechanisms are used to prevent address resolution attacks.
- Request packets can be sent in multicast mode, reducing load on Layer 2 networks.

During address resolution, Neighbor Solicitation (NS) packets and Neighbor Advertisement (NA) packets are used.

- In NS packets, the Type field value is 135 and the Code field value is 0. NS packets are similar to IPv4 ARP Request packets.
- In NA packets, the Type field value is 136 and the Code field value is 0. NA packets are similar to IPv4 ARP Reply packets.

The address resolution process is shown in the following figure:



Host A needs to parse the link-layer address of Host B before sending packets to Host B. Host A sends an NS message with its IPv6 address as the source address and the solicited-node multicast address of Host B as the destination address. The Options field in the NS message carries the link-layer address of Host A.

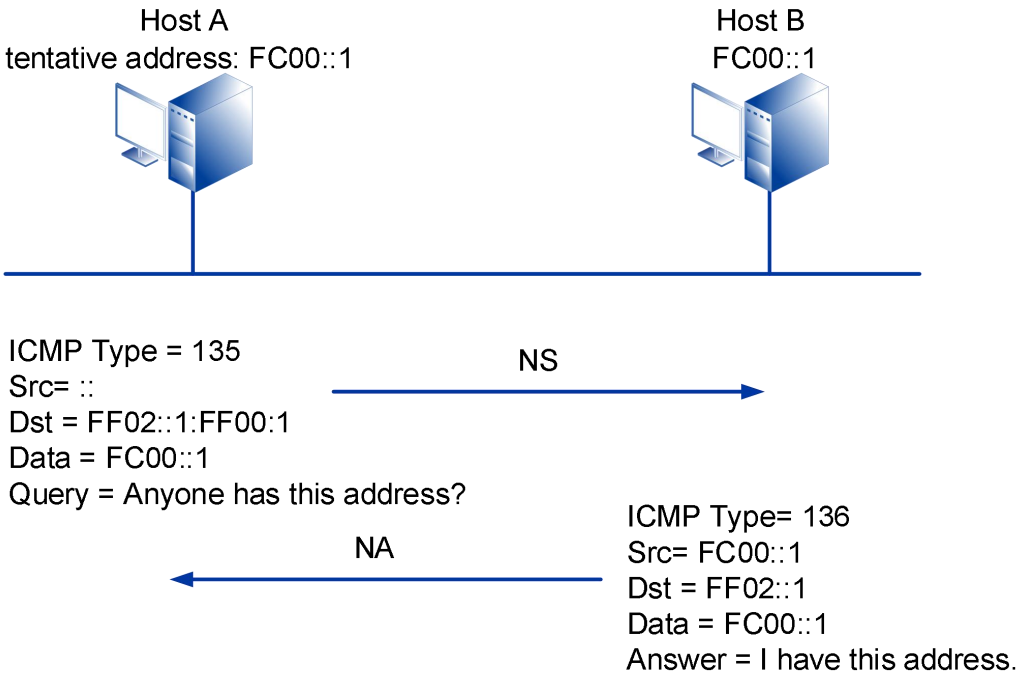
After receiving the NS message, Host B replies with an NA Reply message. In the NA reply message, the source address is the IPv6 address of Host B, and the destination address is the IPv6 address of Host A (the NS message is sent to Host A in unicast mode using the link-layer address of Host A). The Options field carries the link-layer address of Host B. This is the whole address resolution process.

19.2.4.2 Duplicate address detection

Before an IPv6 unicast address is assigned to an interface, duplicate address detection (DAD) is performed to check whether another node uses the address. DAD is required if IP addresses are configured automatically. An IPv6 unicast address assigned to an interface but not verified by DAD is called a tentative address. An interface cannot use the tentative address for unicast communication but will join two multicast groups: ALL-nodes multicast group and Solicited-node multicast group.

IPv6 DAD is similar to IPv4 gratuitous ARP. A node sends an NS message that requests the tentative address as the destination address to the Solicited-node multicast group. If the node receives an NA Reply message, another node is using the tentative address for communication. This node will not use this tentative address for communication.

The principle of duplicate address detection is shown in the following figure:



The IPv6 address FC00::1 is assigned to Host A as a tentative IPv6 address. To check the validity of this address, Host A sends an NS message containing the requested address FC00::1 to the Solicited-node multicast group to which FC00::1 belongs. Since FC00::1 is not specified, the source address of the NS message is an unspecified address. After receiving the NS message, Host B processes the message in one of the following ways:

- If FC00::1 is a tentative address of Host B, Host B will not use this address as an interface address and will not send an NA message.
- If FC00::1 is in use on Host B, Host B sends an NA message to FF02::1 carrying IP address FC00::1. In this way, when Host A receives this message, it will find that its test address is duplicate. The address of the test on Host A is invalid and is marked as duplicated.

19.2.4.3 Router Discovery

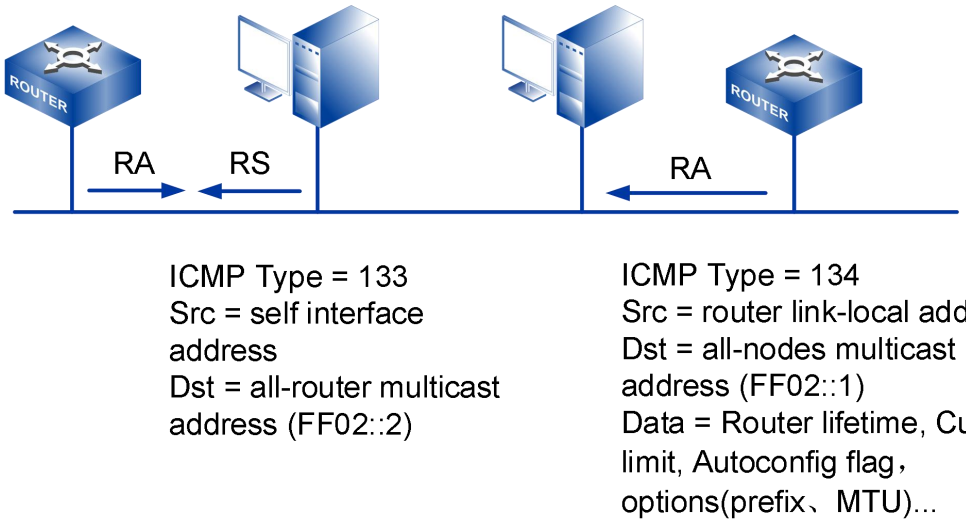
Router discovery is used to locate neighboring devices and learn their address prefixes and configuration parameters for address autoconfiguration.

IPv6 supports stateless address autoconfiguration. Hosts obtain IPv6 prefixes and automatically generate interface IDs. Router Discovery is the basis of IPv6 address autoconfiguration and is implemented through the following two types of packets:

- Router Advertisement (RA) message: Each router periodically sends multicast RA messages carrying network prefixes and identifiers on the network to declare its existence to Layer 2 hosts and devices. An RA message has a Type field value of 134.

- Router Solicitation (RS) message: After being connected to the network, a host immediately sends an RS message to obtain network prefixes. Devices on the network reply with RA messages. An RS message has a Type field value of 133.

The router discovery function is shown in the following figure:



Automatic address configuration

IPv4 uses DHCP to realize automatic configuration, including IP address, default gateway and other information, which simplifies network management. IPv6 addresses grow to 128 bits, and there are many terminal nodes, which makes the automatic configuration more urgent. Besides retaining DHCP as stateful automatic configuration, stateless automatic configuration is also added. Stateless automatic configuration means that the link local address is automatically generated, and the host automatically configures the global unicast address according to the prefix information of RA message, and obtains other relevant information.

The process of IPv6 stateless autoconfiguration is as follows:

1 A host automatically configures the link-local address based on the interface ID.

2 The host sends an NS message for duplicate address detection.

3 If address conflict occurs, the host stops address autoconfiguration. Then addresses need to be configured manually.

4 If addresses do not conflict, the link-local address takes effect. The host then connects to the network and communicates with the local node.

5 The host either sends an RS message or receives RA messages devices periodically send.

6 The host obtains the IPv6 address based on the prefixes carried in the RA message and the interface ID.

Default Router Priority and Route Information Discovery

If there are multiple devices on the network where hosts reside, hosts need to select forwarding devices based on the destination address of the packet. In such a case, devices advertise default router priorities and route information, which allows hosts to select the optimal forwarding device based on the packet destination address.

The fields of default router priority and route information are defined in an RA message. These two fields enable hosts to select the optimal forwarding device.

After receiving an RA message containing route information, hosts update their routing tables. When sending packets to other devices, hosts check the routing table and select the optimal route.

When receiving an RA message carrying default router priorities, hosts update their default router lists. When sending packets to other devices, hosts select the device with the highest priority to forward packets from the router list. If the selected router does not work, hosts select the subsequent device in descending order of priority.

19.2.5 Static Routing

Routing is the most basic element in data communication network. Routing information is the path information guiding message sending, and the routing process is the process of message forwarding.

According to different routing destinations, routes can be divided into:

- Network segment route: The destination is a network segment. In this case, if the destination is an IPv4 address, the subnet mask is less than 32 bits, and if the destination is an IPv6 address, the prefix length is less than 128 bits.
- Host route: The destination is a host. In this case, if the destination is an IPv4 address, the subnet mask is 32 bits, and if the destination is an IPv6 address, the prefix length is 128 bits.

According to whether the destination directly connects to a router, routes are classified into one of the following types:

- Direct route: The router directly connects to the network where the destination is located.
- Indirect route: The router indirectly connects to the network where the destination is located.

According to the destination address type, routes are classified into one of the following types:

- Unicast route: The destination address is a unicast address.
- Multicast route: The destination address is a multicast address.

On the Internet, network connecting devices such as hubs, bridges, switches, and routers control traffic and ensure data transmission quality. Each of these devices serves a different role, but for a common purpose: forming a functioning network.

A router selects routes and forwards packets. Upon receiving a packet, a router selects a proper path, which may have one or multiple hops, to send the packet to the next router according to the destination address in the packet. The last router is responsible for sending the packet to the destination host.

A route is a path along which packets are sent from the source to the destination. When multiple routes are available to send packets from a router to the destination, the router can select the optimal route from an IP routing table. Optimal route selection depends on routing protocol preferences and metrics of routes. When multiple routes have the same routing protocol preference and metric, load balancing can be implemented among these routes to relieve network pressure. When multiple routes have different routing protocol preferences and metrics, route backup can be implemented among these routes to improve network reliability.

Routing protocols are the rules used by routers to discover routes, generate routing tables, and guide packet forwarding. Routes are classified into the following types according to their origin:

- Direct routes: are discovered by link layer protocols.
- Static routes: are manually configured by network administrators.
- Dynamic routes: are discovered by dynamic routing protocols. Dynamic routes include Routing Information Protocol (RIP) routes, Open Shortest Path First (OSPF) routes, and Border Gateway Protocol (BGP) routes.

Static routing configuration is convenient, low requirements on the system, suitable for simple and stable topology of small networks. The disadvantage of static routes is that they require subsequent maintenance as they cannot automatically adapt to network topology changes.

In contrast to dynamic routing, static routing is easier to configure, has higher controllability, uses less bandwidth, and does not use CPU resources for route calculation and update analysis. When a network fault occurs or the topology changes, static routes cannot be automatically updated and must be manually reconfigured to adapt to the network change. Therefore, static routes are not suitable for large and complex networks. In addition, it is difficult for network administrators to know the entire network topology. When the network topology or link state changes, a large amount of static routing information of routers needs to be adjusted, which is a laborious.

19.2.5.1 Destination Address and Mask

The IPv4 destination address of a static route is expressed in dotted decimal notation. The mask of the route can be expressed either in dotted decimal notation or by the mask length. The mask length is the number of consecutive 1s in the mask. An IPv6 Destination address is 128 bits long and is written as eight groups of four hexadecimal digits (base 16 digits represented by the numbers 0-9 and the letters A-F). Each group is separated by a colon (:). IPv6 is composed of network prefix and interface identification, and the network prefix is equivalent to the network id in IPv4 address. Interface ID is equivalent to the host ID in IPv4 address.

Setting the destination and mask to all 0s configures a default static route. The default route is a special route with all zero destination addresses, which can be automatically generated by the routing protocol or manually configured. Manually configuring the default route can simplify the network configuration, which is called static default route. If the destination address of the message cannot match any item in the routing table, the switch will select the default route to forward the message.

19.2.5.2 Egress interface and next hop address

When configuring a static route, depending on the outbound interface type, you need to specify either an outbound interface or a next-hop IP address.

- For point-to-point (P2P) interfaces, specify an outbound interface. This automatically sets the IP address of the remote interface connected to the outbound interface as the next-hop address.
- For non-broadcast multiple access (NBMA) interfaces such as Asynchronous Transfer Mode (ATM) interfaces, specify a next-hop IP address. This type of interface supports point-to-multipoint (P2MP) networks, which require mappings between IP addresses and link-layer addresses to be configured. Therefore, during the configuration of static routes, only a next-hop IP address needs to be specified, and no outbound interface needs to be specified.
- For broadcast interfaces (such as Ethernet interfaces) and virtual template (VT) interfaces, specify a next-hop IP address. Ethernet interfaces are broadcast interfaces, and VT interfaces can be associated with several virtual access (VA) interfaces. If an Ethernet interface or a VT interface is specified as the outbound interface, there will be multiple next hops, and the system will not be able to decide which next hop to use.

19.3 IPv6 Configuration

19.3.1 Create Layer 3 Interface

【Command】

ip interface vlan <VLAN-ID>

【View】

Global configuration mode

【Parameter】

<VLAN-ID> : create a layer 3 vlan interface, the range is 2-4094.

【Description】

ip interface vlan: the command is used to create the specified layer 3 vlan interface.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#vlan database
Switch(config-vlan)#vlan 2
Switch(config-vlan)#exit
Switch(config)#ip interface vlan 2
```

19.3.2 IPV6 Address

【Command】

ipv6 address X:X::X:X/M **[anycast]**

【View】

Layer 3 Interface View

【Parameter】

X:X::X:X/M：X:X::X:X is IPv6 address, M is mask length

anycast：means anycast.

【Description】

ipv6 address: command is used to configure an ipv6 address for the interface

【Instance】

```
Switch> enable
```

```
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ipv6 address 3ffe:506::1/48  //ordinary
ipv6 address
```

19.3.3 Static IPv6Route

【Command】

```
ipv6 route X:X::X:X/M （X:X::X:X | INTERFACE） [<1-255>]
no ipv6 route X:X::X:X/M （X:X::X:X | INTERFACE） [<1-255>]
```

【View】

Global configuration mode

【Parameter】

X:X::X:X/M: specify the static route destination network segment

X:X::X:X: specify ipv6 address of static routing next hop

INTERFACE: outgoing interface for specifying static routing

<1-255> : specifies the value of 1.-255, ranging from 1 to 255.

【Description】

ipv6 route: command is used to add static IPv6 routing.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ipv6 route ::/0 fe80::200:ff:fe00:a0a0 vlanif2
```

19.3.4 Configure RA Message Related Parameters

19.3.4.1 Suppress the publication of RA messages

【Command】

```
ipv6 nd suppress-ra
no ipv6 nd suppress-ra
```

【View】

VLANIF interface view

【Parameter】

None

【Description】

ipv6 nd suppress-RA: suppresses the release of RA messages.

no ipv6 nd suppress-RA: suppresses the release of RA messages.

The RA message is suppressed by default.

When the device is connected with the host and needs to periodically issue the IPv6 address prefix and the information of the stateful automatic configuration flag bit in the RA message to the host, the no ipv6 nd suppress-ra command is used to enable the system to issue the RA message.

When the device is connected with the routing device, that is, when there is no host in the network, it is not necessary to enable the system to issue RA messages, that is, use the default condition of this command.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#no ipv6 nd suppress-ra
```

19.3.4.2 Configure the maximum time interval for RA message publishing

【Command】

ipv6 nd ra-interval <4-1800>

no ipv6 nd ra-interval

【View】

VLANIF interface view

【Parameter】

< 4-1800 >: the maximum time interval for RA message publishing, with a value range of 4-1800, in seconds.

【Description】

ipv6 nd ra-interval: configure the maximum time interval for RA message publishing.

no ipv6 nd ra-interval: the maximum time interval for resuming the publication of ra message is the default value.

By default, the maximum time interval for issuing RA messages is 600 seconds, and the minimum time interval is 198 seconds. When RA messages are issued periodically, the time interval between two adjacent times is to randomly select a value

between the maximum time interval and the minimum time interval as the time interval for periodically issuing RA messages.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#ipv6 nd ra-interval 100
```

19.3.4.3 Configure the minimum time interval for RA message publishing

【Command】

```
ipv6 nd minimum-ra-interval <3-1350>
no ipv6 nd minimum-ra-interval
```

【View】

VLANIF interface view

【Parameter】

< 3-1350 >: the minimum time interval for ra message publishing, with a value range of 3-1350, in seconds.

【Description】

`Ipv6 nd minimum-RA-interval:` configure the minimum time interval for issuing RA messages**. Configure that the minimum interval should be less than or equal to 0.75 times of the maximum interval**

`no ipv6 nd minimum-RA-interval:` the minimum time interval for resuming publication of ra message is the default value.

By default, the maximum time interval for issuing RA messages is 600 seconds, and the minimum time interval is 198 seconds. When RA messages are issued periodically, the time interval between two adjacent times is to randomly select a value between the maximum time interval and the minimum time interval as the time interval for periodically issuing RA messages.

Routing equipment periodically issues RA messages, which include IPv6 address prefixes and information of stateful automatic configuration flags. If you need to change the frequency of RA messages issued by routing equipment, you can use this command. When users need to reduce RA messages on the link, they can use a larger time interval; When users need to speed up router discovery, they can do so by setting a smaller time interval.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#ipv6 nd minimum-ra-interval 75
```

19.3.4.4 Configure the hop limit

【Command】

```
ipv6 nd current-hoplimit <0-255>
no ipv6 nd current-hoplimit [<0-255>]
```

【View】

VLANIF interface view

【Parameter】

< 0-255 >: limit value of hop count, ranging from 0-255.

【Description】

ipv6 nd current-hoplimit: configure the hop limit published by the router.

no ipv6 nd current-hoplimit: restore the hop limit published by the router to the default value.

By default, the number of hops published by routers is limited to 64 hops.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#ipv6 nd current-hoplimit 75
```

19.3.4.5 Configure prefix information of RA messages

【Command】

```
ipv6 nd prefix {X:X::X:X/M | no-autoconf | offlink |
preferred-lifetime <0-4294967295> | valid-lifetime
<0-4294967295>}
no ipv6 nd prefix {X:X::X:X/M | no-autoconf | offlink |
preferred-lifetime <0-4294967295> | valid-lifetime
<0-4294967295>}
```

【View】

VLANIF interface view

【Parameter】

X: x:: x: x/m: IPv6 address and prefix.

< 0-4294967295 >: preferred lifetime or effective lifetime time, with the value range of 0-4294967295, in seconds.

【Description】

ipv6 nd prefix: configure prefix information in RA message.

- X:X::X:X/M: specifies the IPv6 address contained in RA message and the prefix length of IPv6 address.
- No-autoconf: remove the A-Flag flag bit. If this parameter is specified, it means that the configured prefix cannot be used for stateless address autoconfiguration. The A-Flag flag bit is the autonomous address configuration flag bit in the ra message prefix option.
- Offlink: specifies the O-Flag flag bit. If this parameter is configured, the host in this link will be informed that the prefix in the RA message is not assigned to the local link.
- Preferred-lifetime: specifies the preferred lifetime of prefix information. Automatically configured by stateless address during this time. Preferred survival time cannot be greater than effective survival time.
- Valid-lifetime: specifies the valid lifetime of prefix information. The effective time-to-live is used to determine the on-link status of the prefix.

no ipv6 nd prefix: deletes prefix information configured in RA message.

The prefix information in the RA message is not configured by default. In this case, the IPv6 address of the interface sending the RA message will be used as the prefix information in the RA message. Its effective life span is 2592000 seconds (30 days), and the preferred life span is 604800 (7 days).

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#ipv6 nd current-hoplimit 75
```

19.3.4.6 Configure that the MTU option is carried in the RA message

【Command】

```
ipv6 nd link-mtu
no ipv6 nd link-mtu
```

【View】

VLANIF interface view

【Parameter】

None

【Description】

ipv6 nd link-mtu: configure MTU option carried in RA message.

no ipv6 nd link-mtu: the MTU option is not carried in the configuRAtion ra message.

By default, the MTU option is not carried in the RA message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#no ipv6 nd link-mtu
```

19.3.4.7 Set the managed address configuration flag bit

【Command】

ipv6 nd managed-config-flag

no ipv6 nd managed-config-flag

【View】

VLANIF interface view

【Parameter】

None

【Description】

ipv6 nd managed-config-flag: set the managed address configuration flag bit in RA message to 1, then the host obtains IPv6 address through stateful automatic configuration.

no ipv6 nd managed-config-flag: Clear the managed address configuration flag bit in RA message, and the host obtains IPv6 address through stateless automatic configuration. **That is, IPv6 address prefix information is released to the host through RA message to automatically generate IPv6 address.**

By default, the managed address flag bit is 0, that is, the host obtains the IPv6 address through stateless automatic configuration.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#ipv6 nd managed-config-flag
```

19.3.4.8 Set other configuration flags

【Command】

ipv6 nd other-config-flag

no ipv6 nd other-config-flag

【View】

VLANIF interface view

【Parameter】

None

【Description】

ipv6 nd other-config-flag: if other configuRAtion flags in ra message are set to 1, the host can obtain other configuration information except IPv6 address through stateful automatic configuration, including router lifetime, neighbor reachable time, neighbor retransmission time and MTU information of link.

no IPv6 and other-config-flag: Clear other configuration flags in RA message, and then the host will perform stateless automatic configuration, that is, issue other configuration information except IPv6 address to the host through RA message, including router lifetime, neighbor reachable time, neighbor retransmission time and MTU information of link.

By default, the other configuration flag bit is 0, that is, the host obtains other information through stateless automatic configuration.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#ipv6 nd other-config-flag
```

19.3.4.9 Configure the neighbor request message retransmission interval

【Command】

ipv6 nd retransmission-time <1000-3600000>

no ipv6 nd retransmission-time [<1000-3600000>]

【View】

VLANIF interface view

【Parameter】

< 1000-3600000 >: retransmission time, ranging from 1000-3600000, in milliseconds.

【Description】

ipv6 nd retransmission-time: configure the retransmission time interval of neighbor request message sent by the system.Configure the neighbor request message retransmission interval

no ipv6 nd retransmission-time: the retransmission time interval of neighbor request message sent by the recovery system is the default value.

By default, the time interval for the interface to send NS messages is 1000 milliseconds. The value of Retrans Timer field in RA message issued by interface is 0, that is, no host is specified.

Setting the time interval for the system to send neighbor request messages can:

- Control the time interval of neighbor reachability detection for this routing device.
- Control the time interval of repeated address detection for this routing device.
- As a parameter of RA message, inform the host. The host sets this value as the time interval for itself to send neighbor request messages.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#ipv6 nd retransmission-time 1000
```

19.3.4.10 Configure the time to keep neighbors reachable

【Command】

ipv6 nd reachable-time <0-3600000>

no ipv6 nd reachable-time

【View】

VLANIF interface view

【Parameter】

< 0-3600000 >: reaction time, value range 0-3600000, in millisecond.

【Description】

ipv6 nd reachable-time: configure the time to keep the neighbor reachable.

no ipv6 nd reachable-time: the time to restore the neighbor reachable state is the default value.

By default, the interface keeps its neighbor reachable for 30,000 milliseconds. The value of Reachable Timer field in RA message issued by interface is 0, that is, no host is specified.

Every time a router advertisement message is issued on an interface, a routing device carries the neighbor reachable time, so that all nodes on the same link use the same time. Setting the neighbor reachable time of IPv6 neighbor nodes can:

- Controlling the aging time of neighbor entries of the routing equipment;
- As a parameter of RA message, it helps host to configure neighbor reachable time.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif 1
Switch(config-vlanif1)#ipv6 nd reachable-time 30000
```

19.3.5 The Maximum Transmission Unit

【Command】

mtu <MTU-VALUE>

【View】

Layer 3 Interface View

【Parameter】

MTU-VALUE: configure the maximum transmission unit of message, the range is 128-1500, the unit is byte.

【Description】

mtu: command is used to configure the maximum transmission unit of a layer 3 interface message.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#mtu 1400
```

20 NAT Configuration

20.1 Overview

NAT (Network Address Translation) is a process of translating an IP address in an IP data header into another IP address.

With the development of Internet and the increase of network applications, the exhaustion of IPv4 addresses has become a bottleneck restricting the development of network. Although IPv6 can fundamentally solve the problem of insufficient IPv4 address space, many network devices and applications are based on IPv4 at present. Therefore, before IPv6 is widely used, the use of some transitional technologies (such as CIDR, private network address, etc.) is the most important technical means to solve this problem. NAT is mainly used to realize the function that internal network (referred to as internal network, using private IP address) accesses external network (referred to as external network, using public IP address). When the host in the intranet wants to access the external network, the private network address can be converted into the public network address through NAT technology, which can realize that multiple private network users share one public network address to access the external network, thus ensuring network intercommunication and saving the public network address.

As a transitional scheme to slow down the exhaustion of IP addresses, NAT meets the needs of IP addresses through address reuse, which can alleviate the pressure of exhaustion of IP address space to a certain extent. In addition to solving the problem of IP address shortage, NAT also brings two benefits:

- It can effectively avoid attacks from external networks and improve network security to a great extent.
- It can control the internal network host to access the external network, and it can also control the external network host to access the internal network, thus solving the problem that the internal network and the external network cannot

communicate with each other.

20.2 NAT Configuration

20.2.1 Configure NAT rule

【Command】

nat NAT_NAME protocol (tcp|udp|all) outside A.B.C.D inside A.B.C.D dst A.B.C.D/M

no nat NAT_NAME protocol (tcp|udp|all) outside A.B.C.D inside A.B.C.D dst A.B.C.D/M

【View】

Global configuration mode

【Default Level】

2. System level

【Parameter】

NAT_name: the name of the NAT rule, which cannot exceed 20 characters in length.

tcp: the protocol type allowed by NAT is TCP.

udp: the protocol type allowed by NAT is UDP.

all： The allowed protocol types of NAT are tcp, udp and icmp.

A.B.C.D: the converted external network IP.

A.B.C.D: intranet IP before conversion.

A.B.C.D/M: Destination network converted from NAT message.

【Description】

nat NAT_NAME protocol (tcp|udp|all) outside A.B.C.D inside A.B.C.D dst A.B.C.D/M： the command is used to configure NAT rules for non-protocol port number translation.

no nat NAT_NAME protocol (tcp|udp|all) outside A.B.C.D inside A.B.C.D dst A.B.C.D/M： the command is used to delete NAT rules for non-protocol port number translation.

【Instance】

```
# configure a NAT rule named nat-example for no protocol port translation
Switch> enable
Switch#configure terminal
```

```
Switch(config)#nat nat-example protocol all outside 192.168.10.5
inside 192.168.1.5 dst 13.14.15.16/26
```

20.2.2 Configure NAT Destination Network

【Command】

```
nat NAT_NAME add dst A.B.C.D/M
nat NAT_NAME del dst A.B.C.D/M
```

【View】

Global configuration view

【Default Level】

2: System level

【Parameter】

NAT_name: the name of the NAT rule, which cannot exceed 20 characters in length.

A.B.C.D/M: Flexibly add or delete the destination network after NAT message conversion

【Description】

nat NAT_NAME add dst A.B.C.D/M : the command is used to flexibly add the destination network after NAT conversion;

nat NAT_NAME del dst A.B.C.D/M: this command is used to flexibly delete the destination network after NAT translation.

Note:

When adding a destination network segment, it is necessary to ensure that NAT entries already exist;

When deleting a destination network segment, it is necessary to ensure that the deleted destination network segment is not the last one bound to the rule.

【Instance】

```
# add a destination network under a NAT named aaa
Switch> enable
Switch#configure terminal
Switch(config)#nat aaa add dst 192.168.20.0/24
```

20.2.3 Configure NAT Binding Port

【Command】

nat bind NAT_NAME outside IFNAME inside IFNAME

no nat bind NAT_NAME outside IFNAME inside IFNAME

【View】

Global configuration view

【Default Level】

2: System level

【Parameter】

NAT_NAME: the name of the NAT rule, which cannot exceed 20 characters in length.

IFNAME: the extranet port of NAT binding.

IFNAME: the intranet port of NAT binding.

【Description】

nat bind NAT_NAME outside IFNAME inside IFNAME: the command is used to set the binding port of NAT;

no nat bind NAT_NAME outside IFNAME inside IFNAME: the command is used to delete the binding port of NAT.

【Instance】

#configure a NAT port binding named example

Switch> **enable**

Switch#configure terminal

Switch(config)#nat bind example outside vlanif10 inside vlanif2

20.2.4 Delete NAT Rules and Bound Ports

【Command】

no nat NAT_NAME

【View】

Enable view

【Default Level】

2: System level

【Parameter】

NAT_NAME: the name of the NAT rule, which cannot exceed 20 characters in length.

【Description】

no nat NAT_NAME: the command deletes the rules and binding ports of NAT according to its name.

【Instance】

#configure a NAT port binding named example

Switch> **enable**

Switch#configure terminal

Switch(config)#**no nat example**

20.2.5 Display NAT Configuration

【Command】

show nat

【View】

Enable view

【Default Level】

2: System level

【Parameter】

None: Display the basic NAT Configuration Information

【Description】

show nat: The command is used to display NAT configuration in the form of a list.

【Instance】

#configure a NAT port binding named example

Switch> **enable**

Switch#show nat

20.2.6 Display Activated NAT ACL Configuration

【Command】

show nat acl

【View】

Global configuration view

【Default Level】

2: System level

【Parameter】

-

【Description】

show nat acl Command is used to display activated NAT ACL configuration in the form of a chart.

【Instance】

#configure a NAT port binding named example

Switch> **enable**

Switch# show nat acl

21 DHCP Configuration

21.1 Overview

DHCP (Dynamic Host Configuration Protocol) is a technology used for centralized and dynamic user IP address management and configuration. Even in small networks, DHCP is useful because it makes it easy to add new devices to the network.

DHCP is based on the Bootstrap Protocol (BOOTP), which runs in a static environment where each client has a fixed network connection. For each client using BOOTP, a network administrator must configure a BOOTP parameter file that requires manual intervention to modify. DHCP extends BOOTP from the following two aspects:

- DHCP allows computers to obtain IP addresses dynamically, instead of statically assigning addresses to each host.
- You can also use DHCP to deliver configuration parameters, such as a configuration file used for startup, to clients.

DHCP is defined in RFC 2131 and enables the automatic configuration of DHCP clients. It removes the need to configure clients individually and consists of two components: a protocol for delivering client-specific configuration parameters from a DHCP server to a client, and a mechanism for allocating network addresses to clients.

DHCP can provide two address allocation mechanisms, and network administrators can choose different allocation strategies for different hosts according to network requirements.

- Dynamic allocation: DHCP assigns an IP address to a client for a limited period (or until the client releases the address).

  This mechanism allows automatic reuse of an IP address that is no longer needed by the client to which it was assigned. It is useful for assigning an IP address to a client that connects to the network only temporarily or for sharing a limited pool of IP addresses among a group of clients that do not need permanent

IP addresses.

- Static allocation: A network administrator assigns an IP address to a client and uses DHCP to deliver this address to the client.
  This mechanism allows you to eliminate the error-prone process of manually configuring IP addresses for clients.

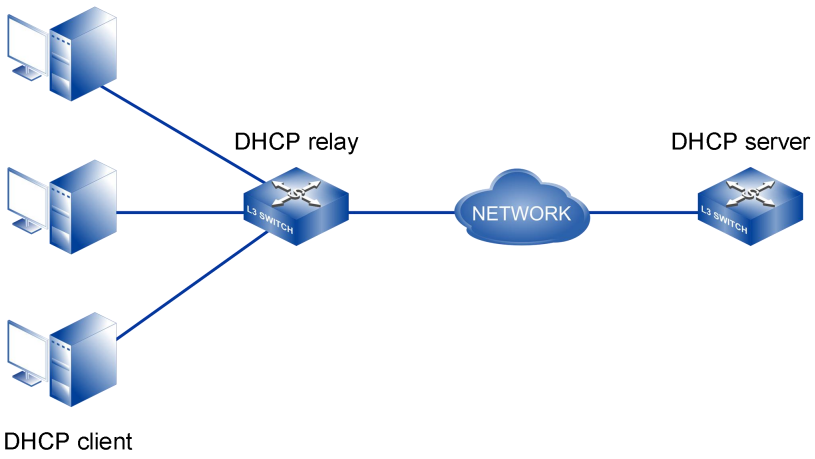
DHCP offers the following benefits:

- Reduced client configurations and costs
  Because DHCP is easy to configure, it minimizes operational costs associated with device configurations, eases deployment by non-technical users, and reduces device configuration and maintenance costs at remote sites.
- centralized management
  Because the DHCP server maintains configurations for several subnets, an administrator only needs to update a single, central server when configuration parameters change.

21.2 Principle Description

21.2.1 Network Elements in DHCP

Typical DHCP networking is shown in the following figure.



The following describes the three roles involved in DHCP:

- DHCP Server
  A DHCP server assigns IP addresses from specified address pools to DHCP clients. It can also manage these clients and provide network parameters such as the default gateway address, Domain Name System (DNS) server address, and Windows Internet Name Service (WINS) server address. A DHCP server can accept broadcasts from locally attached LAN segments or DHCP requests

forwarded by DHCP relay agents within the network.

- DHCP client
  A client can use BOOTP or DHCP to obtain its IP address and other network parameters from a DHCP server. To obtain an IP address, the client sends a BOOTP or DHCP Request message. DHCP clients can be IP phones, PCs, mobile devices, diskless workstations, or other networked devices, and can be connected directly or through other networks using DHCP relay agents.
- DHCP Relay
  DHCP relay agent forwards DHCP messages between a DHCP server and DHCP clients and helps the DHCP server to dynamically allocate network parameters to the DHCP clients.
  When a DHCP client broadcasts DHCP Discovery messages with the destination IP address 255.255.255.255, only the DHCP server on the same network segment as the DHCP client can receive the messages. If a DHCP server is on a different network segment from the DHCP client, a DHCP relay agent must be deployed to forward DHCP Discovery messages to the DHCP server. The DHCP relay agent modifies the format of a DHCP Discovery or Offer message to generate a new DHCP message and then forwards it.

21.2.2 DHCP Leases and Address Pools

21.2.2.1 Lease term

A lease is defined as the time period for which a DHCP server allocates an IP address to a client. The lease can be extended upon subsequent requests. If the client no longer needs the IP address, it can release the address back to the server before the lease expires. The server is then free to assign this address to a different client if no other idle IP address is available.

The lease period configured for a DHCP server applies to all of the IP addresses that a DHCP server dynamically assigns to its clients. A different DHCP server may have a different lease term for its clients. A statically allocated IP address is not subject to the lease terms.

A DHCP client does not wait for its lease to expire, because it may be assigned a different IP address. Instead, when a DHCP client reaches the halfway point of its lease period, it attempts to extend its lease so that it retains the same IP address.

21.2.2.2 Address pool

An address pool is a set of all the IP addresses that a DHCP server has reserved for dynamic client allocation. In addition to IP address, network parameters such as lease period, subnet mask and default gateway can also be configured in the address pool. When DHCP server assigns IP addresses to clients, these network parameters are also assigned to clients.

Address pools are classified into interface address pools and global address pools.

- Interface address pool: After an IP address is configured for an interface on a DHCP server, you can create an address pool on the same network segment as this interface. Addresses in the address pool can be allocated only to clients connected to the interface. The interface address pool can allocate IP addresses to clients on the same network segment as the DHCP server.
- Global address pool: On a DHCP server, you can create an address pool on the specified network segment in the system view. Addresses in the address pool can be allocated to all clients connected to the DHCP server, even if the server and clients are on different network segments (providing that a DHCP relay agent is used).

A DHCP server selects address pools according to whether a DHCP relay agent is deployed. When no relay agent is deployed, the server selects the address pool on the same network segment as the IP address of the interface receiving DHCP Request messages. When relay agents are deployed, the server selects the address pool on the same network segment as the IP address specified in the giaddr field of received DHCP Request messages.

The number of IP addresses required in an address pool depends mainly on the number of clients that will connect to the network and the frequency at which they connect and disconnect.

IP addresses in an address pool can be in the following status based on the IP address usage:

- Used: indicates that this IP address is already in use.
- Idle: indicates that this IP address is idle.
- Static-bind: indicates that this IP address is bound with MAC address and is not used.
- Static-bind used: indicates that this IP address is bound to the MAC address and has been used.
- Disable: indicates that this IP address cannot be used.
- Expired: indicates that the lease term of this IP address has expired and is in an idle state.

  After an IP address in an address pool expires, it is in Expired status. Allocation

records of IP addresses in Expired status are retained, so that when a user requests an IP address again, the previously associated IP address can be directly allocated to the user, ensuring stability of user IP addresses.

When IP addresses in Idle status are exhausted, the address pool automatically reclaims the IP addresses in Expired status and allocate the IP addresses to the users without the need to manually clear the IP addresses.

- Conflict: indicates that this IP address conflicts with other addresses on the network.

  When an IP address in Conflict status exists in an address pool, an IP address conflict is prevented in advance. An IP address in Conflict status will exist in the following situations:

  - When a DHCP server receives a DHCP Discover message from a client, it sends a ping packet before allocating an IP address to the client. If the ping operation succeeds, the server sets the IP address status to Conflict and allocates another IP address to the client.
  - After the DHCP client successfully obtains an IP address, it immediately sends a gratuitous ARP packet. If a response packet is received, the client sends a DHCP Decline message to the DHCP server to notify the DHCP server that the IP address is in conflict. The DHCP server then sets the IP address status to Conflict, and the client sends a DHCP Discover message to request for an IP address again.

  When IP addresses in Idle and Expired status in an address pool are exhausted, the address pool automatically reclaims the IP addresses in Conflict status. The server then allocates the IP addresses to new users without the need to manually clear the address pool.

21.2.3 DHCP Messages

21.2.3.1 DHCP Message Types

A DHCP server and a DHCP client communicate by exchanging DHCP messages. DHCP messages are transmitted using the User Datagram Protocol (UDP). A DHCP client uses UDP port 68 to send messages to a DHCP server, and a DHCP server uses UDP port 67 to send messages to a DHCP client. These messages are classified into eight types.

DHCP Message	Note

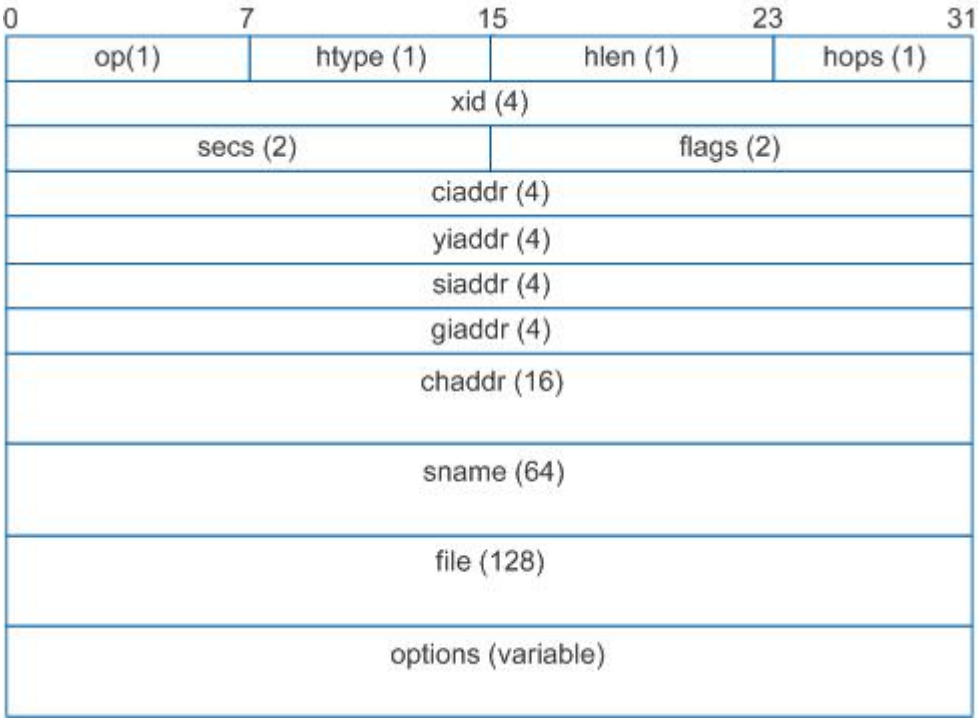
DHCP Message	Note
DHCP Discover	A DHCP client broadcasts this message to locate a DHCP server when the client attempts to connect to a network for the first time.
DHCP Offer	A DHCP server sends this message in response to a DHCP Discover message. A DHCP Offer message carries configuration information.
DHCP Request	A DHCP client sends this message in the following scenarios: • After the client starts, it broadcasts a DHCP Request message to respond to a DHCP Offer message sent by a DHCP server. • After the client restarts, it broadcasts a DHCP Request message to confirm the configuration (including the allocated IP address). • After the client obtains an IP address, it unicasts or broadcasts a DHCP Request message to renew the IP address lease.
DHCP Ack	A DHCP server sends this message to acknowledge a DHCP Request message sent from a DHCP client. After receiving a DHCP Ack message, the DHCP client obtains configuration parameters (including an IP address).
DHCP Nak	A DHCP server sends this message to reject a DHCP Request message from a DHCP client. For example, a DHCP server will send this message if it determines that there is no available IP address after receiving a DHCP Request message.
DHCP Decline	A DHCP client sends this message to notify the DHCP server that the allocated IP address conflicts with another IP address. The DHCP client then applies to the DHCP server for another IP address.
DHCP Release	A DHCP client sends this message to release its allocated IP address. After receiving a DHCP Release message, the DHCP server can allocate this IP address to another DHCP client.
DHCP Inform	A DHCP client sends this message to obtain network configuration parameters, such as the gateway address and

DHCP Message	Note
	DNS server address, after it has obtained an IP address.

21.2.3.2 DHCP Message Format

The format of DHCP messages is based on the format of BOOTP messages, which ensures support for BOOTP functionality and interoperability between BOOTP clients and DHCP servers.

The format of DHCP message is shown in the following figure, and the number in brackets indicates the length of field, and the unit is byte.



The meaning of each field in DHCP message is shown in the following table.

Field	Length	Definition
op (op code)	1 bytes	Represent the type of message, the value is 1 or 2, meaning is as follows: • 1: DHCP Discover message • 2: DHCP Offer message
htype (hardware type)	1 bytes	Hardware Type: indicates the type of hardware used for the local network. The

Field	Length	Definition
		values of this field differ for different hardware types. The most common value is 1, which indicates Ethernet (10 Mb).
hlen (hardware length)	1 bytes	Hardware Address Length: indicates the length of a hardware address. For Ethernet, the value is 6.
hops	1 bytes	Hops: indicates the number of DHCP relay agents through which a DHCP message passes. This value is set to 0 by a client and is incremented by 1 each time the message passes through a DHCP relay agent. A DHCP message passes through a maximum of 16 DHCP relay agents when being transmitted between a server and a client. That is, the number of hops between the server and client cannot exceed 16. Otherwise, the DHCP message is discarded.
xid	4 bytes	Transaction Identifier: indicates a random number chosen by a client for exchanging messages with a DHCP server.
secs (seconds)	2 bytes	Seconds: indicates the number of seconds elapsed since a client obtained or renewed an IP address.
flags	2 bytes	Flags: indicates the Flags field. Only the leftmost bit in this field is used, and the other bits are set to 0. The leftmost bit specifies the mode a DHCP server uses to transmit a DHCP Offer message. • 0: The DHCP server unicasts a DHCP Offer message. • 1: The DHCP server broadcasts a DHCP Offer message.
ciaddr	4 bytes	Client IP Address: indicates the IP address

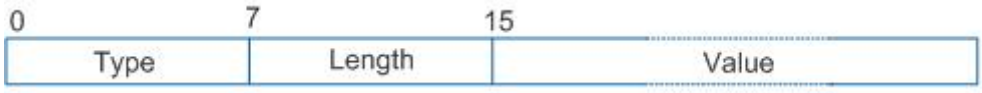
Field	Length	Definition
(client ip address)		of a DHCP client. The IP address is either the existing IP address of the client or an IP address allocated by a DHCP server to the client. During the process of a client acquiring an IP address, the value of this field is 0.0.0.0. 0.0.0.0 is an invalid destination address and is used by a DHCP-enabled device to communicate only temporarily with other devices during startup.
yiaddr (your client ip address)	4 bytes	Indicates the IP address assigned by the server to the client. When the server makes DHCP response, fill in this field with the IP address assigned to the client.
siaddr (server ip address)	4 bytes	Server IP Address: indicates the server IP address from which a DHCP client obtains its startup configuration file.
giaddr (gateway ip address)	4 bytes	Gateway Address: indicates the IP address of the first DHCP relay agent. When a client sends a DHCP Request message and is on a different network segment from its DHCP server, the first DHCP relay agent forwards the message to the DHCP server and fills its IP address in the giaddr field. The DHCP server determines the network segment address of the client based on this field, selects an appropriate address pool, and assigns an IP address on this network segment to the client. The server also returns a DHCP reply message to the first DHCP relay agent, which then forwards the message to the client. If the DHCP Request message passes

Field	Length	Definition
		through multiple DHCP relay agents before reaching the DHCP server, the giaddr field value is still the IP address of the first DHCP relay agent, and the hops field value is incremented by 1 each time the message passes through a DHCP relay agent.
chaddr (client hardware address)	16 bytes	Indicates the MAC address of the client. This field is consistent with the previous "hardware type" and "hardware length". When the client sends a DHCP request, fill in this field with its own hardware address. For Ethernet, this field must contain a 6-byte Ethernet MAC address if the hardware type and hardware length fields are set to 1 and 6 respectively.
sname (server host name)	64 bytes	Server Hostname: indicates the name of the server from which a client obtains configuration information. This field is filled in by DHCP server and is optional. If filled in, it must be a string ending in 0.
file (file name)	128 bytes	Boot File: indicates the name of the startup configuration file to be obtained by a client. This field is filled in by a DHCP server and delivered to the client when a DHCP address is allocated to the client. The field is optional and must be a character string that ends with 0.
options	Variable	Options: indicates the DHCP Options field. This field is a maximum of 312 bytes in length and contains the DHCP message type and configuration parameters allocated by a DHCP server to a client. The configuration parameters include the gateway IP address,

Field	Length	Definition
		DNS server IP address, and IP address lease.

21.2.3.3 DHCP Options

DHCP Options are tagged data items that provide control information and parameters to a DHCP client. The options are sent in a variable-length field at the end of a DHCP message. As shown in the following figure, the Options field consists of three parts: Type, Length and Value. The meanings of these three parts are shown in the following table.



Field	Length	Definition
Type	1 bytes	Indicates the information type.
Length	1 bytes	Indicates the length of the subsequent content in the Options field.
Value	Variable	Indicates the message content. The length varies depending on the Length field.

The values of the Options field range from 1 to 255. Some DHCP options are predefined and others can be user defined. Introduce some predefined options of DHCP Options as shown in the following table.

Option Code	Function
1	Specifies a subnet mask.
3	Specifies a gateway address.
6	Specifies the IP address of a DNS server.
12	Specifies the device name of a DHCP client.

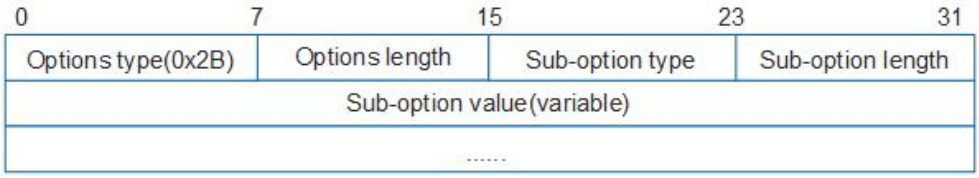
Option Code	Function
15	Specifies a domain name.
33	Specifies a group of classful static routes that the DHCP client must add to its routing table. This option contains a group of classified static routes (that is, the mask of destination address is fixed as natural mask, and subnet cannot be divided). After receiving this option, the client will add these static routes to the routing table. If Option 121 is configured, Option 33 is ignored.
44	Specifies a NetBIOS server name.
46	Specifies a NetBIOS node type.
50	Specifies a requested IP address.
51	Specifies an IP address lease.
52	Specifies an additional option.
53	Specifies a DHCP message type.
54	Specifies a server identifier.
55	Specifies a parameter request list. The client uses this option to indicate which network configuration parameters it needs to get from the server. The option content is the option value corresponding to the parameter requested by the client.
58	Specifies the lease renewal time (T1), which is 50% of the lease time.
59	Set the renewal T2 time. Generally, it is 87.5% of the lease time.
60	Specifies the vendor category, which identifies the DHCP client type and configuration.
61	Specifies a client identifier.
66	Specifies a TFTP server name allocated to DHCP clients.
67	Set boot filename option to specify the boot filename assigned to the client.
77	Specifies a user type.

Option Code	Function
121	Specifies a group of classless static routes that the DHCP client must add to its routing table. This option contains a set of unclassified static routes (that is, the destination address's mask is an arbitrary value that can be used to divide the subnet) that are added to the routing table when the client receives the option. Note: A device functioning as a DHCP client can receive static routes delivered from a DHCP server through Option 121.

In addition to the predefined options, you can configure user-defined options to support a wide variety of devices, such as IP phones.

- Vendor-specific information option (Option 43)

  Option 43 is called the vendor-specific information option. The message format of Option43 is shown in the figure below：



  DHCP servers and DHCP clients exchange vendor-specific information through Option 43. When a DHCP server receives a DHCP Discover message with parameter 43 encapsulated in Option 55, it encapsulates Option 43 in a DHCP Offer message and sends the message to the DHCP client.

- Relay agent information option (Option 82)

  Option 82 is the DHCP relay agent information option that records the location of a DHCP client. A DHCP relay agent or a device with DHCP snooping enabled appends the Option 82 field to a DHCP Discover message sent from a DHCP client and then forwards the DHCP Discover message to a DHCP server.

  The administrator can use the Option 82 field to locate a DHCP client and control the security and accounting of the DHCP client. The DHCP server that supports the Option 82 field can determine policies to allocate IP addresses and other parameters according to information in the Option 82 field. IP addresses can be allocated flexibly.

  The Option 82 field contains a maximum of 254 sub-options. If the Option 82 field is defined, at least one sub-option must be defined. Sub-options that are currently supported by the device include:

- sub-option1 (Agent Circuit ID Sub-option)
- sub-option2 (Agent Remote ID Sub-option)

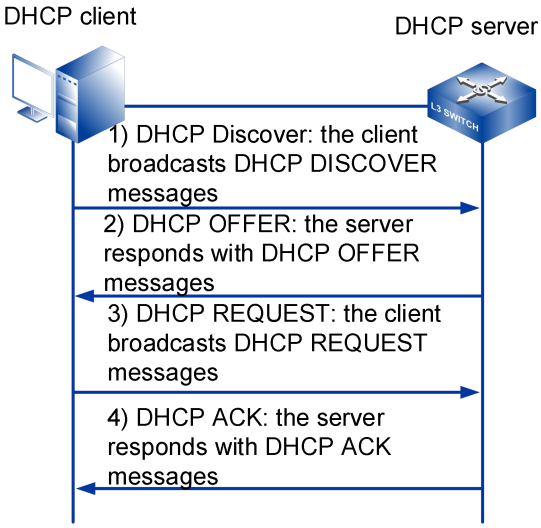
21.2.4 The DHCP Server Assigns a Network Address to the Client That Accesses for the First Time

When a DHCP client broadcasts DHCP Discover messages, only a DHCP server on the same network segment as the DHCP client can receive the messages. If the DHCP client is on a different network segment from the DHCP server, a DHCP relay agent must be deployed to forward DHCP messages between the DHCP client and server.

Depending on whether a DHCP relay agent is used, the way in which network parameters are allocated to a new DHCP client differs.

21.2.4.1 Network Parameter Allocation Without a DHCP Relay Agent

Figure shows the message exchange process between a DHCP server and a new DHCP client when no DHCP relay agent is deployed. This process is called four-message exchange.



Stage 1: The Discovery Stage

When a DHCP client accesses a network for the first time, it does not know the IP address of the DHCP server. To learn this information, the client broadcasts a DHCP Discover message in which the destination IP address is 255.255.255.255 to all devices (including the DHCP server or relay agent) on the network segment. The

DHCP DISCOVER message includes the client's MAC address (chaddr field), parameter request list (Option 55), and broadcast flag (flags field).

Stage 2: The Offer Stage

The DHCP server on the same network segment as the DHCP client receives the DHCP Discover message, selects an available IP address from the address pool that is on the same network segment as the IP address of the interface that receives the DHCP Discover message, and then sends a DHCP Offer message carrying the selected IP address to the DHCP client.

In most cases, an address pool specifies the lease of the IP addresses it contains. If the DHCP Discover message contains an expected lease, the server compares the expected lease with the specified lease and allocates an IP address with the shorter of the two leases to the client.

The DHCP server selects an IP address for a client from the address pool in the following sequence:

1 IP address statically bound to the MAC address of the client on the DHCP server
2 Address specified by Option55 (request IP address option) in DHCP DISCOVER message sent by client.
3 IP addresses in Expired status in the address pool, that is, the allocated IP addresses whose lease time expires
4 Look for an IP address in "Idle" status randomly in the address pool.
5 If no available IP address is found, the address pool automatically reclaims the IP addresses in Expired and Conflict status. If an IP address is available after the reclaim, the server allocates this IP address. Otherwise, the DHCP client sends a DHCP Discover message again to request an IP address after the timeout interval for the client to wait for a response from the server expires.

You can specify certain IP addresses to exclude on the DHCP server. For example, if you have statically allocated 192.168.1.100/24 to your DNS server, you can exclude this IP address from the address pool on network segment 192.168.1.0/24 so that it is not allocated through DHCP. This helps prevent IP address conflicts.

To prevent a newly allocated IP address conflicting with existing IP addresses, the DHCP server sends an ICMP Echo Request packet before sending a DHCP Offer message. This ICMP packet contains the IP address to be allocated in both the source and destination IP address fields. The server can allocate the IP address if it receives no ICMP Echo Reply packet within the detection period (no client is using this IP address). If the server receives an ICMP Echo Reply packet within the detection period, the DHCP server lists this IP address as a conflicting IP address (as

it is in use by another client), and then waits for the next DHCP Discover message to start the IP address selection process again.

The IP address allocated during the offer stage may not be the final IP address used by the client. This is because the IP address may be allocated to another client if the DHCP server receives no response 16 seconds after the DHCP Offer message is sent. The IP address for the client can be determined only after the request and acknowledgment stages.

Stage 3: The Request Stage

If multiple DHCP servers respond to DHCP OFFER messages to the DHCP client, the DHCP client generally only receives the first received DHCP OFFER message, and then sends a DHCP REQUEST message by broadcasting, which contains the DHCP server identifier (Option54) that the client wants to select and the IP address of the client (Option50, which is filled with the IP address of yiaddr field in the received DHCP OFFER message).

The DHCP Request message notifies all the DHCP servers of the IP address that the DHCP client has selected. The unselected IP addresses offered by other DHCP servers are then free to be allocated to other clients.

Stage 4: The Acknowledgment Stage

Upon receiving the DHCP REQUEST message sent by the DHCP client, the DHCP server responds to the DHCP ACK message, indicating that the IP address requested in the DHCP REQUEST message (filled with Option50) is assigned to the client for use.

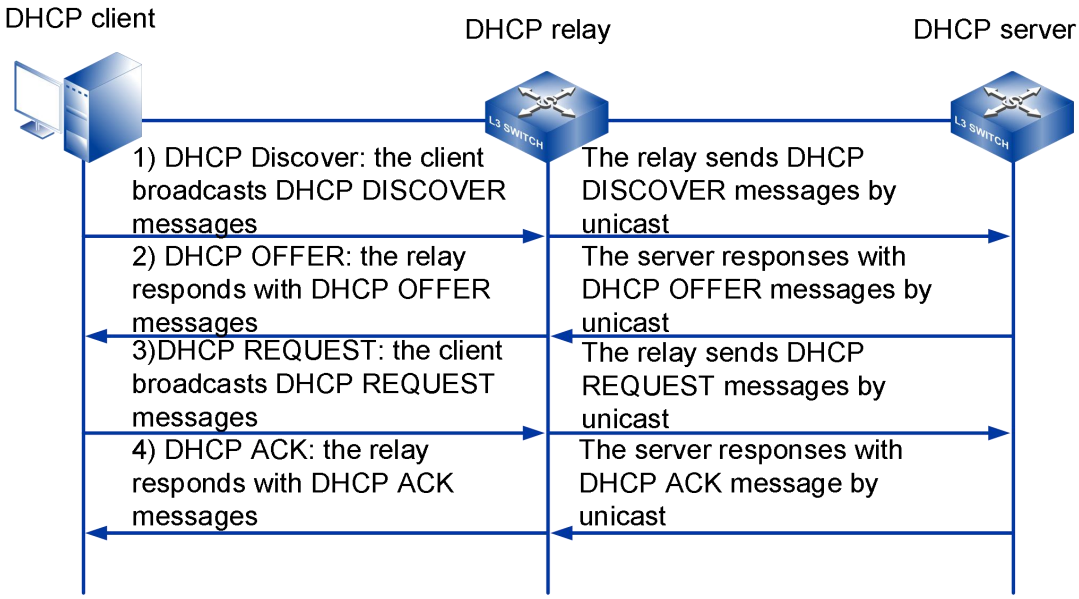
To determine whether any other device is using this IP address, the DHCP client broadcasts gratuitous ARP packets after receiving the DHCP Ack message. The client can use this IP address if no response is received within the specified time. However, if the DHCP client receives a response within the specified time, this IP address is already in use. The client then sends a DHCP Decline message to the DHCP server and applies for a new IP address. The server lists this IP address as a conflicting IP address.

Occasionally, the DHCP server may not allocate the IP address specified in the Option 50 field because, for example, an error occurs during negotiation or it does not receive the DHCP Request message quickly enough. In this case, the server replies with a DHCP NAK message to notify the client that the requested IP address cannot be allocated. The client then sends a DHCP Discover message to apply for a new IP address.

21.2.4.2 Network Parameter Allocation with a DHCP Relay Agent

The message exchange process between a DHCP server and a new DHCP client when a DHCP relay agent is deployed is similar to that described in Network Parameter Allocation without a DHCP Relay Agent. The main difference is that the DHCP relay agent acts as an intermediary to forward DHCP messages between a DHCP server and client that would otherwise be unable to communicate with each other. The following describes how the DHCP relay agent functions in the message exchange process.

Figure shows the message exchange process between a DHCP server and a new DHCP client when a DHCP relay agent is deployed.



Stage 1: The Discovery Stage

When a DHCP relay agent receives a DHCP Discover message, it performs the following steps:

1 Check the value of the hops field. If this value exceeds 16, the relay agent discards the message. Otherwise, the relay agent increases this value by 1 and proceeds to the next step.

2 Check the value of the giaddr field. If this value is 0, the relay agent sets the giaddr field to the IP address of the interface receiving the DHCP Discover message. Otherwise, the relay agent does not change the field and proceeds to the next step.

3 Change the destination IP address of the DHCP Discover message to the IP address of the DHCP server or the next-hop DHCP relay agent, and change the source IP address to the IP address of the interface connecting the DHCP relay

agent to the client. The relay agent then unicasts this message to the DHCP server or the next-hop DHCP relay agent.

If there are multiple DHCP relay agents between the DHCP client and server, each the DHCP relay agent processes the DHCP Discover message using the same method.

Stage 2: The Offer Stage

After receiving a DHCP Discover message, the DHCP server selects an address pool on the same network segment as that specified in the giaddr field and allocates an IP address and other network parameters from the address pool. The DHCP server then unicasts a DHCP Offer message to the DHCP relay agent specified in the giaddr field. When the DHCP relay agent receives a DHCP Offer message, it performs the following steps:

1 Check the value of the giaddr field. If this value is the IP address of the interface receiving the DHCP Offer message, the DHCP relay agent discards the message. Otherwise, the relay agent proceeds to the next step.

2 Check the value of the flags field. If this value is 1, the DHCP relay agent broadcasts a DHCP Offer message to the DHCP client. Otherwise, the DHCP relay agent unicasts a DHCP Offer message.

Stage 3: The Request Stage

The DHCP relay agent processes the DHCP Request message from the client in the same manner as that described in Stage 3: The Request Stage.

Stage 4: The Acknowledgment Stage

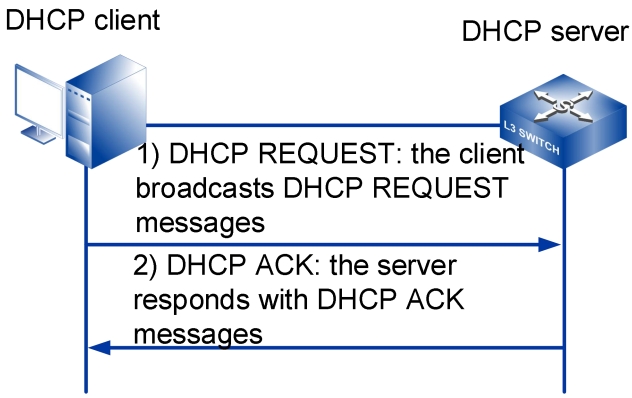
The DHCP relay agent processes the DHCP Ack message from the server in the same manner as that described in Stage 4: The Acknowledgment Stage.

21.2.5 DHCP Client Reuses Network Address

If a DHCP client reconnects to a network, it may be able to reuse the IP address that it had been previously allocated. The DHCP client exchanges DHCP messages with a DHCP server to attempt to obtain the previously used network parameters, including the IP address. Figure shows this message exchange, which is called two-message exchange.

Note

Not all clients can reuse IP addresses that have been allocated to them.



Stage 1: The Request Stage

The DHCP client broadcasts a DHCP Request message that contains the IP address used previously by the client. The requested IP address is added in the Option 50 field.

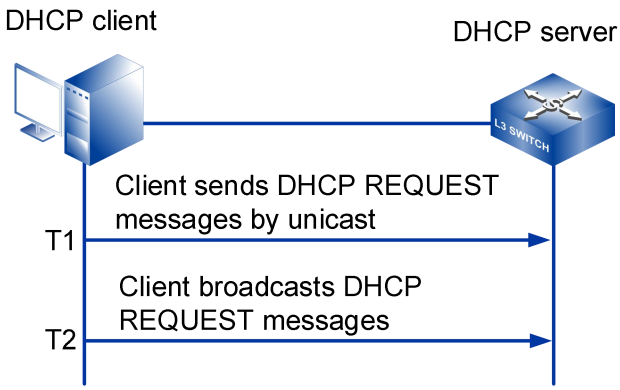
Stage 2: The Acknowledgment Stage

After receiving the DHCP Request message, the DHCP server checks whether there is a lease record based on the MAC address in the message. If there is a lease record matching the MAC address, the DHCP server replies with a DHCP Ack message to notify the client that it can use the requested IP address. Otherwise, the server ignores the request and waits for a new DHCP Discover message from the client.

21.2.6 DHCP Client Renews Its IP Address Lease

IP addresses that are dynamically allocated by a DHCP server have leases. However, a DHCP client may request a specific lease term by adding information to a DHCP Discover message. When allocating network parameters, the DHCP server compares the expected lease with the lease specified in the address pool and allocates an IP address with a smaller lease to the DHCP client. When the lease expires or a client logs out and releases its IP address, the server reclaims the IP address, which can then be allocated to other clients. To continue using this IP address, the client must renew its IP address lease.

Figure shows how a DHCP client renews its IP address lease.



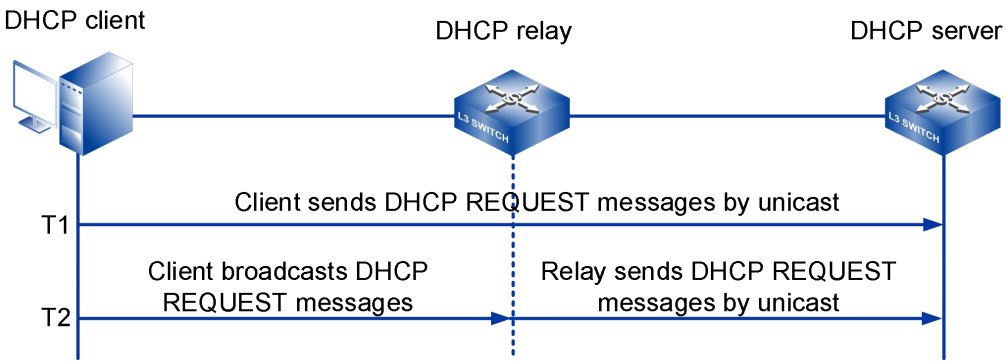
1 When the lease reaches 50% (T1) of its validity period, the DHCP client unicasts a DHCP Request message to the DHCP server to request lease renewal. If the server renews the lease (counted from 0), it sends a DHCP Ack message to the client. If the server rejects the renewal request, it sends a DHCP Nak message to the client, which must then send a DHCP Discover message to apply for a new IP address.

2 If no response is received from the DHCP server when the lease reaches 87.5% (T2) of its validity period, the DHCP client broadcasts a DHCP Request message to request lease renewal. If the server renews the lease (counted from 0), it sends a DHCP Ack message to the client. If the server rejects the renewal request, it sends a DHCP Nak message to the client, which must then send a DHCP Discover message to apply for a new IP address.

3 If no response is received when the lease expires, the client stops using the IP address and sends a DHCP Discover message to apply for a new IP address.

When a DHCP client no longer needs to use its allocated IP address and the lease has not expired, the client sends a DHCP Release message to the DHCP server to request IP address release. The server saves the configuration of this client and records the IP address in the allocated IP address list. The IP address can then be allocated to this client or other clients. To request a configuration update, a client can send a DHCP Inform message to the server.

The renewal process is similar when a DHCP relay agent is used. Figure shows how a DHCP client renews its IP address lease when a DHCP relay agent is deployed.



21.3 DHCP Configuration

21.3.1 Global DHCP Service Enablement

【Command】

ip dhcp service
no ip dhcp service

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

No.

【Description】

ip dhcp service: command is used to globally enable the dhcp server service.

【Instance】

```
#Configure to enabled DHCP server service globally
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
```

21.3.2 Enable Interface DHCP Relay

【Command】

ip dhcp relay enable
no ip dhcp relay enable

【View】

Layer 3 Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

No.

【Description】

ip dhcp relay enable: the command is used to enable the dhcp relay function of the interface.

no ip dhcp relay enable: the command is used to disable the dhcp relay function of the interface.

【Instance】

```
# Enable dhcp relay on vlanif1
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip dhcp relay enable

# disable DHCP relay on vlanif1
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#no ip dhcp relay enable
```

21.3.3 InterfaceDHCP Relay Server Address

【Command】

```
ip dhcp relay-to <A.B.C.D>
no ip dhcp relay-to [<A.B.C.D>]
```

【View】

Layer 3 Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: dhcp server IP address.

【Description】

This command is in the Interface view and only configure the corresponding Interface relay and dhcp server IP addresses.

ip dhcp relay-to <A.B.C.D>: the command is used to set the dhcp server ip address required by the relay. This command can be executed repeatedly to configure multiple server IP addresses, which up to 8 relay servers can be configured.

no ip dhcp relay-to [<A.B.C.D>]: used to delete a relay server ip address, the function is opposite to ip dhcp relay-to <A.B.C.D>; With no parameter, the command deletes all relay server ip addresses of the corresponding interface.

【Instance】

```
Configure the dhcp relay server IP address for vlanif 1,the parameter IP address is 192.168.1.1
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip dhcp relay enable
Switch(config-vlanif1)#ip dhcp relay-to 192.168.1.1

# Remove the DHCP relay server address with IP of 192.168.1.1 under vlanif 1
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#no ip dhcp relay-to 192.168.1.1

# Remove all the dhcp relay server address under vlanif 1
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#no ip dhcp relay-to
```

21.3.4 DHCP Option82 Enablement

【Command】

ip dhcp option

no ip dhcp option

【View】

Layer 3 Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

-

【Description】

Ip dhcp option: enable the option 82 function of dhcp relay, and enable the relay message sent by the relay process to carry option 82.

no ip dhcp option: disable option 82 function of dhcp relay.

【Instance】

```
For vlanif1 interface, enable option 82 function of dhcp relay
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip dhcp relay enable
Switch(config-vlanif1)#ip dhcp option
For vlanif1 interface, enable option 82 function of dhcp relay
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#no ip dhcp option
```

21.3.5 Treatment Strategy of DHCP Option82

【Command】

ip dhcp relay-information policy （append | discard | replace | untouched ）

no ip dhcp relay-information policy

【View】

Layer 3 Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

append: configure the option check policy as Append.

discard: configure the option check policy as Discard.

replace: configure the option check policy as Replace.

untouched: configure the option check policy as Untouched.

【Description】

ip dhcp relay-information policy (append | discard | replace | untouched): set an option processing policy of dhcp relay, the relay process will process the received dhcp message with option 82 according to the policy.

no ip dhcp relay - information policy: disable the option processing strategy of dhcp relay, strategy will restore to the default strategy, namely only forward received dhcp message with the option 82 (Untouched).

【Instance】

```
# configure option 82 with the policy replace for vlanif1 port.
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip dhcp relay enable
Switch(config-vlanif1)#ip dhcp option
Switch(config-vlanif1)#ip dhcp relay-to 192.168.1.1
Switch(config-vlanif1)#ip dhcp relay-information policy replace

#Disable option 82 policy on port vlanif1.
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#no ip dhcp relay-information policy
```

21.3.6 Relay Identity of DHCP Option82

【Command】

ip dhcp relay-information circuitid (basic | string OPTION)

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

Basic: configure sub option circuited is the base (default) configuration.

String: configure sub option circuited as the string given by option.

【Description】

ip dhcp relay-information circuitid (basic | string OPTION): Set the value of option82 suboption circuitid of DHCP relay.

【Instance】

```
#Configure the value of suboption circuited of option 82 is the string "vlan2:ge10” for vlanif1 port.
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip dhcp relay enable
Switch(config-vlanif1)#ip dhcp option
Switch(config-vlanif1)#ip dhcp relay-information circuitid string vlan2:ge10
```

21.3.7 Remote Identity of DHCP Option82

【Command】

ip dhcp relay-information remoteid (basic | string OPTION)

【View】

Layer 3 Interface Configuration View

【Default Level】

2: Configuration level

【Parameter】

Basic: configure sub option remoteid is the base (default) configuration.

String: configure sub option remoteid as the string given by option.

【Description】

ip dhcp relay-information remoteid (basic | string OPTION): Set the value of option82 suboption remoteid of DHCP relay.

【Instance】

```
Configure the value of sub option circuited of option 82 is the string "00:11:22:33:44:55” for vlanif1 port.
Switch> enable
Switch#configure terminal
Switch(config)#interface vlanif1
Switch(config-vlanif1)#ip dhcp relay enable
Switch(config-vlanif1)#ip dhcp option
Switch(config-vlanif1)#ip dhcp relay-information remoteid string 00:11:22:33:44:55
```

21.3.8 Create DHCP Address Pool

【Command】

ip dhcp pool <WORD>

no ip dhcp pool <WORD>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

WORD: the name of DHCP address pool.

【Description】

ip dhcp pool: the command is used to build the DHCP address pool.

no ip dhcp pool: the command is used to delete the DHCP address pool.

【Instance】

```
#Create a dhcp address pool named "test" :
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#

#Delete a dhcp address pool named "test" :
Switch> enable
Switch#configure terminal
Switch(config)#no ip dhcp pool test
```

21.3.9 DHCP Address Pool Subnet Segment

【Command】

network (<A.B.C.D A.B.C.D> | <A.B.C.D/M>)

no network

【View】

DHCP Configuration View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: IP address & mask, used to represent subnet segments.

A.B.C.D/M: IP address and mask,used to represent subnet segments.

【Description】

Network:: configure the subnet segment of DHCP pool , among which the parameters give the specific subnet segment value.

no network: deletes the subnet segment configuration of a specified DHCP pool.

【Instance】

```
#Configure network for address pool "test”
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#network 192.168.1.1 255.255.255.0
Switch(config-dhcppool_test)#network 192.168.1.1/24

#Delete network configuration of address pool "test”
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#no network
```

21.3.10 Default Route of DHCP Address Pool

【Command】

default-router A.B.C.D

no default-router

【View】

dhcp View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: The default routing address used by dhcp.

【Description】

default-router: configure the default routing IP address of DHCP pool.

no default-router: delete the default routing configuration of DHCP pool.

【Instance】

```
#Configure the default route to 192.168.1.1 of dhcp pool test
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#network 192.168.1.1 255.255.255.0
Switch(config-dhcppool_test)#default-router 192.168.1.1

#delete the default routing configuration of dhcp pool test
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#no default-router 192.168.1.1
```

21.3.11 DHCP Address Pool

【Command】

```
range  <A.B.C.D A.B.C.D>
no range (<A.B.C.D A.B.C.D>|)
```

【View】

dhcp View

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D　A.B.C.D: The lowest and highest addresses of dhcp pool.

【Description】

range: configure the address range of DHCP pool, that is, the addresses that belong to the range can be allocated effectively by DHCP.

no range: delete the address range of DHCP pool. If no range parameters are given, delete all range configurations.

【Instance】

```
Configuration the address range of dhcp pool test to: 192.168.1.10
192.168.1.20
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
```

```
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#network 192.168.1.1 255.255.255.0
Switch(config-dhcppool_test)#range 192.168.1.10 192.168.1.20

Delete the address range of dhcp pool test: 192.168.1.10
192.168.1.20
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#no range 192.168.1.10 192.168.1.20

Delete all the address range of dhcp pool test:
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#no range
```

21.3.12 The Lease Time of DHCP Address Pool

【Command】

lease-time <0-30> <0-24> <0-60>

no lease-time

【View】

dhcp View

【Default Level】

2: Configuration level

【Parameter】

<0-30> : days.

<0-24> : hours.

<0-60> : minutes.

【Description】

lease -time: configure the address lease duration of dhcp pool. When the IP address obtained by the dhcp client is about to reach the lease duration, it is necessary to renew the lease. Otherwise, the IP address will be invalid, and the dhcp client needs to re-request the IP address.

no lease-time: delete the address lease duration configuration of dhcp pool and restore the lease duration to the default value.

【Instance】

```
#set the lease-time of dhcp pool test to 30 minutes.
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#network 192.168.1.1 255.255.255.0
Switch(config-dhcppool_test)#lease-time 0 0 30

# set the lease-time of dhcp pool test.
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
Switch(config)#ip dhcp pool test
Switch(config-dhcppool_test)#no lease-time
```

21.3.13 DNS Server Address

【Command】

ip dhcp dns-server <A.B.C.D> [<A.B.C.D>] [<A.B.C.D>]

no ip dhcp dns-server

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

< A.B.C.D>: dns server ip address.

【Description】

ip dhcp dns-server: configure dns servers for all dhcp pools, up to three different dns servers can be configured (this configuration is global).

no ip dhcp dns-server: delete all dns server configurations.

【Instance】

```
# set the DNS -serve of DHCP pool to 114.114.114.114 8.8.8.8.
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
Switch(config)#ip dhcp dns-server 114.114.114.114 8.8.8.8
```

```
Remove all dns-serve configurations.
Switch> enable
Switch#configure terminal
Switch(config)#no ip dhcp dns-server
```

21.3.14 Log Server Address

【Command】

ip dhcp log-server <A.B.C.D> [<A.B.C.D>] [<A.B.C.D>]
no ip dhcp log-server

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

< A.B.C.D>: log server ip address.

【Description】

ip dhcp log-server: configure log servers for all dhcp pools, up to three different log servers can be configured (this configuration is global).

no ip dhcp log-server: delete all log server configurations.

【Instance】

```
# set the log-server of dhcp pool to 192.168.1.1 192.168.1.2
192.168.1.3.
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
Switch(config)#ip dhcp dns-server 192.168.1.1 192.168.1.2
192.168.1.3

#Remove all log-server configurations.
Switch> enable
Switch#configure terminal
Switch(config)#no ip dhcp log-server
```

21.3.15 WINS Server Address

【Command】

ip dhcp wins-server <A.B.C.D> [<A.B.C.D>] [<A.B.C.D>]

no ip dhcp wins-server

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

< A.B.C.D>: wins server ip address.

【Description】

ip dhcp wins-server: configure wins servers for all dhcp pools, up to three different wins servers can be configured (this configuration is global).

no ip dhcp win-server: delete all winserver configurations.

【Instance】

```
# configure the wins-server of dhcp pool to be 192.168.1.1
192.168.1.2 192.168.1.3.
Switch> enable
Switch#configure terminal
Switch(config)#ip dhcp service
Switch(config)#ip dhcp wins-server 192.168.1.1 192.168.1.2
192.168.1.3

#Delete all wins-server configurations.
Switch> enable
Switch#configure terminal
Switch(config)#no ip dhcp wins-server
```

21.3.16 Display DHCP Information

21.3.16.1 View DHCP Global Information

【Command】

show ip dhcp global

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

-

【Description】

show ip dhcp global: view DHCP global information.

【Instance】

```
Switch#show ip dhcp global
             Global Servers
--------------------------------------------


--------------------------------------------
```

21.3.16.2 View DHCP Lease Information

【Command】

show ip dhcp lease ((interface [IFNAME]) | (summary [IFNAME]))

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

[IFNAME]: interface name.

【Description】

show ip dhcp lease: view DHCP lease information.

【Instance】

```
*Switch#show ip dhcp lease summary

DHCP Lease Information Summary(current time : 1970/01/01 13:25:25 )
Interface       Allocated Lease    Expired Lease     Total Lease
-------------------------------------------------------------
-------
vlanif1                    0              0              0
```

```
lo                          0               0               0
```

21.3.16.3 View DHCP Address Pool Information

【Command】

show ip dhcp pool [WORD]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

WORD: address pool identification.

【Description】

show ip dhcp pool: view DHCP address pool information.

【Instance】

```
Switch#show ip dhcp pool

Pool test :
----------------------
  network: 192.168.1.0/24
  address range(s):
    add: 192.168.1.100 to 192.168.1.102
  lease <days:hours:minutes> <1:0:0>
  no default-routers
```

21.3.16.4 View DHCP Relay Information

【Command】

show ip dhcp relay [IFNAME]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

[IFNAME]: interface name.

【Description】

show ip dhcp relay: view DHCP relay information.

【Instance】

```
Switch#show ip dhcp relay
%Server Lists for DHCP Relay : interface vlanif1, option 0
```

21.3.16.5 View DHCP Statistics Information

【Command】

show ip dhcp statistics [IFNAME]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

[IFNAME]: interface name.

【Description】

show ip dhcp statistics: view DHCP statistics information.

【Instance】

```
Switch#show ip dhcp statistics
DHCP Statistics : vlanif1
------------------------------------------------------------
rxDhcpDiscovers  :      24   txDhcpDiscovers  :       0
rxDhcpRequests   :       0   txDhcpRequests   :       0
rxDhcpOffers     :       0   txDhcpOffers     :       0
rxDhcpAcks       :       0   txDhcpAcks       :       0
rxDhcpNaks       :       0   txDhcpNaks       :       0
rxDhcpDeclines   :       0   txDhcpDeclines   :       0
rxDhcpReleases   :       0   txDhcpReleases   :       0
rxDhcpInforms    :       0   txDhcpInforms    :       0
rxDhcpBadPackets :       0   txErrorPackets   :       0
------------------------------------------------------------
rxTotalPackets   :      24   txTotalPackets   :       0
------------------------------------------------------------
------------------------------------------------------------
rxBootpRequest   :       0   txBootpRequest   :       0
rxBootpReply     :       0   txBootpReply     :       0
------------------------------------------------------------
```

21.3.16.6 View DHCP Status Information

【Command】

show ip dhcp status

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

-

【Description】

show ip dhcp status: view DHCP status information.

【Instance】

```
Switch#show ip dhcp status
Interface       IP Address      DHCP Status
-----------------------------------------------
vlanif1         192.168.1.254   DHCP Relay
-----------------------------------------------
```

22 DHCP-Snooping Configuration

22.1 Overview

DHCP Snooping is a security feature of DHCP (Dynamic Host Configuration Protocol) that ensures that DHCP clients receive IP addresses from valid DHCP servers. The corresponding relationship between IP address of DHCP client and MAC address is recorded to prevent DHCP attack on network.

At present, there are many security problems in the application of DHCP protocol (RFC2131). There are some attacks against DHCP in the network, such as DHCP Server impersonator attack, DHCP Server denial of service attack, DHCP packet impersonation attack and so on. In order to ensure the security of network communication service, DHCP Snooping technology can be introduced to build a firewall between DHCP Client and DHCP Server to resist various attacks against DHCP in the network. The equipment has the ability to defend against DHCP attack on the network, which enhances the reliability of the equipment and ensures the normal operation of the communication network. To provide users with more secure network environment, more stable network services.

22.2 Principles

22.2.1 DHCP Snooping Principles

A DHCP Snooping device will send the DHCP request message from the user (DHCP client) to the valid DHCP server through the trust interface. After that, the device generates the DHCP Snooping binding table according to the DHCP ACK message information responded by the DHCP server. When the subsequent devices receive

DHCP messages from the user through the enabled DHCP Snooping interface, they will carry out a match check to effectively prevent the attack of illegal users.

DHCP Snooping Trust Function

The trust function of DHCP Snooping can ensure that the client gets the IP (Internet Protocol) address from the valid server.

If there is a privately established DHCP Server impersonator in the network, it may cause the DHCP client to obtain the wrong IP address and network configuration parameters, and cannot communicate normally. The DHCP Snooping trust function can control the source of the DHCP Server reply messages to prevent the possible network DHCP Server impersonators from assigning IP addresses and other configuration information to the DHCP clients.

The DHCP Snooping trust function divides the interfaces into trusted and untrusted interfaces:

- • The trust interface normally receives DHCP ACK, DHCP NAK and DHCP offer messages from the DHCP server. In addition, the device will only send DHCP request packets from the DHCP client to the legitimate DHCP server through the trust interface.
- • Untrusted interface discards DHCP ACK, DHCP NAK, and DHCP OFFER packets after receiving the DHCP server response.

In the scenario of DHCP Snooping enabled by Layer 2 network access devices, the interface directly or indirectly connected with the legitimate DHCP server is generally set as trusted interface, and the other interface is set as untrusted interface, so that the DHCP request message of the DHCP client can only be forwarded from the trusted interface. In this way, DHCP clients can only obtain IP addresses from legitimate DHCP servers, and the impersonator of DHCP Server can not assign IP addresses to DHCP clients.

DHCP Snooping Binding Table

The DHCP Snooping binding table will be aged according to the DHCP lease term or the corresponding table entries will be automatically deleted according to the DHCP Release message issued by the user when the IP address is released.

Since the DHCP Snooping binding table records the corresponding relationship between the IP address of the DHCP client and the MAC address, it can effectively prevent the attack of illegal users by checking the match between the message and the DHCP Snooping binding table.

In order to ensure that the device can obtain the user's MAC and other parameters when generating the DHCP Snooping binding table, the DHCP Snooping function

shall be applied to the access devices in the Layer 2 network or on the first DHCP Relay.

In the DHCP Relay enabling DHCP Snooping scenario, the DHCP Relay device does not need to set up the trust interface. Since the DHCP Relay receives the DHCP request message, it will process the IP and MAC conversion of the source destination, and then send it to the designated legitimate DHCP server in the form of unicast, the DHCP ACK messages received by the DHCP Relay are all valid, and the DHCP Snooping binding table generated is also correct.

22.2.2 Option82 Supported by DHCP Snooping

In the traditional DHCP dynamic IP address allocation process, the DHCP Server can not perceive the specific physical location of the user according to the DHCP request packet, so that the user of the same VLAN gets the IP address with the same permissions. Because the network manager can not control the specific user in the same VLAN effectively, that is, can not control the client access to the network resources, which will give a severe challenge to the network security control.

RFC 3046 defines the DHCP Relay Agent Information Option (Option 82), which records the location Information of the DHCP Client. The DHCP Snooping device or DHCP Relay passes the exact physical location of the DHCP Client to the DHCP Server by adding Option82 to the DHCP request message. Thus, DHCP Server can assign appropriate IP address and other configuration information to the host, and realize the security control of the client.

Option82 contains two commonly used sub-options, Circuit ID and Remote ID. The Circuit ID sub-option is mainly used to identify the VLAN, interface and other information where the client is, while the Remote ID sub-option is mainly used to identify the device accessed by the client, which is generally the MAC address of the device.

When the device is DHCP Relay, Option82 can be supported by enabled or unenabled DHCP Snooping. However, if the device is an access device on the Layer 2 network, Option82 can only be supported if DHCP Snooping is enabled.

Option82 only records the exact physical location information of DHCP users and sends the information to DHCP Server through DHCP request message. If you need to deploy different address assignment or security policies for different users, then the DHCP Server needs to support Option82 and have IP address assignment or security policies configured on it.

The user location information carried by the Option82 option and the user parameters recorded by the DHCP Snooping binding table are two separate concepts and have no connection. The user location information carried by Option82 is added by the device to the DHCP request packet when the DHCP user applies for an IP address (the user has not been assigned an IP address at this time). The DHCP Snooping binding form is generated automatically when the device receives the DHCP ACK message responded by the DHCP Server (the user has been assigned an IP address at this time).

Implementation

Option82 can be supported when a device serves as DHCP Relay or when a device serves as an access device on a Layer 2 network and enables DHCP Snooping. The Option82 function of enabled devices has two modes of Insert and Rebuild. Different devices with different enabling modes can process DHCP request packets differently.

- Insert: When the device receives DHCP request packet, if Option82 is not in the text, then Insert Option82; If Option82 is included in this text, determine whether Option82 contains remote-id. If so, keep Option82 unchanged; if not, insert remote-id.
- Rebuild mode: When the device receives DHCP request packet, if there is no Option82 option in this text, insert Option82 option; If Option82 is included in the text, delete it and insert the Option82 configured by the administrator himself on the device.

For Insert and Rebuild, when the device receives the response packet from the DHCP server, it is handled in the same way.

- The DHCP response message has the option Option82:
  - If Option82 is not available in the DHCP request received by the device, the device will delete Option82 from the DHCP response and forward it to the DHCP Client.
  - If Option82 is included in the DHCP request received by the device, the device reverts the format of Option82 in the DHCP response to Option82 in the DHCP request and forwards it to the DHCP Client.
- The DHCP response message does not contain the Option82 option: direct forward.

22.3 Configure DHCP Snooping

22.3.1 Configure the Interface to Trust state

【Command】

ip dhcp snooping trust

no ip dhcp snooping trust

【View】

Port view

【Default Level】

2: Configuration level

【Parameter】

-

【Description】

ip dhcp snooping trust: The command is used to configure the Layer 2 port properties to a new trust.

no ip dhcp snooping trust:The trust property used to cancel the port.

【Instance】

Configure port ge10 as trust

```
Switch> enable
switch# configure terminal
switch(config)# interface ge10
switch(config)# ip dhcp snooping trust
```

Cancel port ge10's trust

```
Switch> enable
switch# configure terminal
switch(config)# interface ge10
switch(config)# no ip dhcp snooping trust
```

22.3.2 Configure the Rate of a DHCP Packet for the Layer 2 Port Interface

【Command】

ip dhcp snooping limit rate <1-2048>

no ip dhcp snooping limit rate

【View】

Port view

【Default Level】

2: Configuration level

【Parameter】

<1-2048> : packet reception rate, PCS /s.

【Description】

ip dhcp snooping limit rate <1-2048>: The command is used to configure the rate of DHCP packets on the Layer 2 port interface. The parameter value is the rate size, in units /s.

no ip dhcp snooping limit rate: The DHCP packet receiving rate limit used to remove the port.

【Instance】

DHCP port GE10 receives the DHCP message at a limit rate of 512

```
Switch> enable
switch# configure terminal
switch(config)# interface ge10
switch(config)# ip dhcp snooping limit rate 512
```

Cancel limit rate on port ge10 to receive DHCP messages

```
Switch> enable
switch# configure terminal
switch(config)# interface ge10
switch(config)# no ip dhcp snooping limit rate
```

22.3.3 Enable Port Disable Duration

【Command】

ip dhcp snooping errdisable recovery cause dhcp-rate-limit
no ip dhcp snooping errdisable recovery cause dhcp-rate-limit

【View】

Global configuration view

【Default Level】

2: Configuration level

【Parameter】

-

【Description】

Enable/disable Settings for the amount of time required to restore a port (port disabled time).

When a port is disabled due to receiving messages exceeding the speed limit, this duration determines the length of time to be disabled, or the length of time required to recover. The default time value is 30s.

【Instance】

```
Switch> enable
switch# configure terminal
switch(config)# ip dhcp snooping errdisable recovery cause dhcp-rate-limit
```

22.3.4 Configure Port Disable Duration

【Command】

```
ip dhcp snooping errdisable recovery interval <1-3600>
no ip dhcp snooping errdisable recovery interval
```

【View】

Global configuration view

【Default Level】

2: Configuration level

【Parameter】

<1-3600> : duration, the unit is second.

【Description】

Sets/deletes the value of the time required to restore the port (port disabled time).

When a port is disabled due to receiving messages exceeding the speed limit, this duration determines the length of time to be disabled, or the length of time required to recover. The default time value is 30s.

【Instance】

Set port overspeed disabling time to 120 seconds

```
Switch> enable
switch# configure terminal
switch(config)# ip dhcp snooping errdisable recovery interval 120
```

22.3.5 Delete DHCP Snooping security database entry

【Command】

```
clear ip dhcp snooping binding (dynamic|A.B.C.D)
```

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

Dynamic: dynamic item.

A.B.C.D: An entry with an IP address of the specified value.

【Description】

Delete a table entry from a secure database, where all dynamic table entries are deleted if the parameter dynamic is given. If the argument is a specific IP address, delete the entry corresponding to that IP.

【Instance】

Delete all the dynamic DHCP snooping safety table entry

```
Switch> enable
switch# clear ip dhcp snooping binding dynamic
```

delete dhcP snooping database entry eith ip 192.168.1.100

```
Switch> enable
switch# clear ip dhcp snooping binding 192.168.1.100
```

22.3.6 DHCP Snooping Enablement

【Command】

```
ip dhcp snooping
no ip dhcp snooping
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

-

【Description】

Enable/disable DHCP snooping function. This command is a global command; DHCP snooping and DHCP Server Relay are in conflict with each other and cannot be enabled at the same time.

【Instance】

Enable dhcp Snooping

```
Switch> enable
switch# configure terminal
switch(config)# ip dhcp snooping
```

disable DHCP snooping function

```
Switch> enable
switch# configure terminal
switch(config)#no ip dhcp snooping
```

22.3.7 VLAN DHCP Snooping Enable

【Command】

```
ip dhcp snooping vlan <1-4096>
no ip dhcp snooping vlan <1-4096>
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<1-4096>: vlan id。

【Description】

Enable/disable DHCP snooping for a VLAN.

【Instance】

Enable vlan 1's dhcp snooping

```
Switch> enable
switch# configure terminal
switch(config)# ip dhcp snooping vlan 1
```

Disable DHCP snooping for VLAN 1

```
Switch> enable
switch# configure terminal
switch(config)#no ip dhcp snooping vlan 1
```

22.3.8 Configure DHCP Snooping Static Binding Entries

【Command】

```
ip dhcp snooping binding vlan <1-4096> interface IFNAME A.B.C.D MAC
no ip dhcp snooping binding vlan <1-4096> interface IFNAME A.B.C.D
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<1-4096> : The bound VLAN ID.

IFNAME: binded interface name

A.B.C.C：binding IP address.

MAC: The address of the bound MAC.

【Description】

When a DHCP snooping static binding is added/removed, the IP address of the binding will be reserved (such as for static configuration). Any conflicting IP address will be rejected by DHCP snooping.

【Instance】

Add a static bounding with ip 192.168.1.100 and mac c2:23:6d:45:8a:f3 in vlan 1, ge10:

```
Switch> enable
switch# configure terminal
switch(config)# ip dhcp snooping binding vlan 1 interface ge10 192.168.1.100 c223-6d45-8af3
```

Delete a static bouding with IP 192.168.1.100 in vlan1, ge10

```
Switch> enable
switch# configure terminal
switch(config)#no ip dhcp snooping binding vlan 1 interface ge10 192.168.1.100
```

22.3.9 Enable DHCP Snooping Check Function for MAC

【Command】

```
ip dhcp snooping verify mac-address
```

```
no ip dhcp snooping verify mac-address
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

-

【Description】

Enable DHCP Snooping check function for MAC

This function is global. When enabled, it will check whether the MAC address of the client in the DHCP message is the same as the MAC address of the header on the layer 2 of the message. If it is different, the message will be discarded.

【Instance】

Enable verify mac-address function for DHCP snooping

```
Switch> enable
switch# configure terminal
switch(config)# ip dhcp snooping verify mac-address
```

Disable verify mac-address function for DHCP snooping

```
Switch> enable
switch# configure terminal
switch(config)#no ip dhcp snooping verify mac-address
```

22.3.10 Enable Option 82 Check for DHCP Snooping

【Command】

```
ip dhcp snooping information
no ip dhcp snooping information
```

【View】

Port view

【Default Level】

2: Configuration level

【Parameter】

-

【Description】

When the Option 82 function of DHCP snooping is enabled, Option 82 check is performed on the DHCP request packet file received by the corresponding port, and the file is processed according to different configured policies.

【Instance】

```
Switch> enable
switch# configure terminal
switch(config)# interface ge10
switch(config-ge10)# ip dhcp snooping information
```

22.3.11 Configure the Option 82 Handling Policy for DHCP Snooping

【Command】

```
ip dhcp snooping information strategy < drop | keep | replace >
```

【View】

Port view

【Default Level】

2: Configuration level

【Parameter】

keep: configure the option check policy as keep.

drop: Configure the Option Check policy to drop.

Replace: configure the option check policy as Replace.

【Description】

Set the Option 82 processing policy for DHCP snooping, according to which the DHCP snooping process will process the DHCP file it receives carrying Option 82.

【Instance】

Enable Option 82 on port GE10 and configure the strategy to replace.

```
Switch> enable
switch# configure terminal
switch(config)# interface vlanif1
switch(config-if)# ip dhcp snooping information
switch(config-if)# ip dhcp snooping information strategy replace
```

22.3.12 Configure the Option 82 Circuit-id of DHCP Snooping

【Command】

```
ip dhcp snooping information circuit-id (hex|ascii) WORD
no ip dhcp snooping information circuit-id
```

【View】

Port view

【Default Level】

2: Configuration level

【Parameter】

Hex: Configure option 82 circuit-id with a string type of hex (hex).

ASCII: Configure option 82 circuit-id with string type ASCII (string).

WORD: Configure the value of option 82 circuit-id.

【Description】

Configure the option 82 circuit-id of DHCP Snooping, can be configured to HEX, ASCII type according to the need.

【Instance】

Enable GE10 port 82 circuit-id with ASCII value "01".

```
Switch> enable
switch# configure terminal
switch(config)# interface ge10
switch(config-if)# ip dhcp snooping information circuit-id ascii
01
```

22.3.13 Configure the Option 82 Remote-id of DHCP Snooping

【Command】

```
ip dhcp snooping information remote-id (hex|asciibasic) WORD
no ip dhcp snooping information remote -id
```

【View】

Port view

【Default Level】

2: Configuration level

【Parameter】

Hex: Configure option 82 Remote-id with a string type of hex (hex).

ASCII: Configure option 82 Remote-id with string type ASCII (string).

The value of option 82 remote-id is the base configuration, which is the device MAC address.

WORD: Configure the option 82 Remote-id of DHCP Snooping

【Description】

Configure the option 82 Remote-id of DHCP Snooping, can be configured to HEX, ASCII type according to the need.

【Instance】

Enable GE10 port 82 remote-id with ASCII value "test".

```
Switch> enable
switch# configure terminal
switch(config)# interface ge10
switch(config-if)# ip dhcp snooping information circuit-id ascii
test
```

22.3.14 Configure the Value of Option 82 Subscriber-id for DHCP Snooping

【Command】

```
ip dhcp snooping information subscriber-id (ascii) WORD
no ip dhcp snooping information subscriber -id
```

【View】

Port view

【Default Level】

2: Configuration level

【Parameter】

ASCII: Configure option 82 subscriber-id with string type ASCII (string).

WORD: Configure the option 82 subscriber-id of DHCP Snooping

【Description】

Configure the option 82 subscriber-id of DHCP Snooping, can be configured to HEX, ASCII type according to the need.

【Instance】

Enable GE10 port 82 subscriber-id with ASCII value "aaaa".

```
Switch> enable
switch# configure terminal
switch(config)# interface ge10
switch(config-if)# ip dhcp snooping information subscriber-id
ascii aaaa
```

22.3.15 Display DHCP Snooping Information

```
show ip dhcp snooping
show ip dhcp snooping （(vlan <1-4096>)|)
show ip dhcp snooping binding
show ip dhcp snooping binding vlan <1-4096> (IFNAME|)
```

23 SNMP Configuration

23.1 Overview

SNMP (Simple Network Management Protocol) is a network management standard protocol widely used in TCP/IP network. SNMP provides a way to manage devices by running network management software on a central computer (or network management workstation). The features of SNMP are as follows:

- Simple: SNMP adopts polling mechanism that provides the most basic function set, and is suitable for small, fast and low-price environment. Moreover, SNMP is supported by most devices because it is carried by UDP messages.
- Powerful: The goal of SNMP is to ensure that management information is transmitted at any two points, so that administrators can retrieve information at any node on the network and troubleshoot.

With the rapid development of network technology, while the network continues to popularize, it also brings some problems to network management:

- The number of network devices increases geometrically, which makes it more and more difficult for network administrators to manage the devices. At the same time, as a complex distributed system, the network covers an expanding area, which makes it extremely difficult to monitor and troubleshoot these devices in real time.
- There are many kinds of network devices, and the management interfaces (such as command line interface) provided by different device manufacturers are different, which makes the network management more and more complex.

Under this background, SNMP came into being. Through "using network to manage network", SNMP realizes efficient and batch management of network devices; At the same time, SNMP protocol also shields the differences between different products, and realizes the unified management among network devices of different kinds and manufacturers.

In May 1990, RFC 1157 defined the first version of SNMP, SNMPv1. RFC 1157 provides a systematic method for monitoring and managing computer networks. SNMPv1 is based on community name authentication, which has poor security and few error codes of returned messages.

Later, IETF issued SNMPv2c. GetBulk and Inform operations are introduced into SNMPv2c, which supports more standard error code information and more data types (Counter64 and Counter32).

In view of the fact that the security of SNMPv2c has not been improved, IETF has issued the version of SNMPv3, which provides authentication encryption based on USM(User-based Security Model) and access control based on VACM (View-based Access Control Model).

Advantages of SNMP (simple network management protocol);

- Network administrators can use SNMP platform to complete information query, information modification and fault troubleshooting on any node on the network, and the work efficiency can be improved.
- The physical differences between devices are shielded, and SNMP only provides the most basic function set, which makes the management tasks independent of the physical characteristics and network types of managed devices, thus realizing unified management of different devices with low management cost.
- The design is simple and the running cost is low. SNMP adopts the design idea of "as simple as possible". The software/hardware added to the devices, the types of messages and the format of messages are all simple, so the impact and cost caused by running SNMP on the devices are minimized.

23.2 Principles

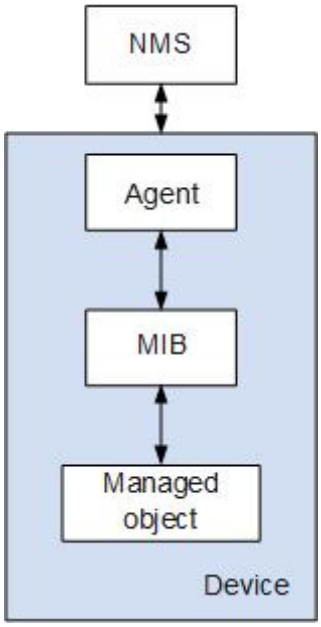
23.2.1 SNMP Management Model

SNMP System consists of NMS (Network Management System), Agent Process, Management Object and MIB (Management Information Base) four parts.

As the network management center of the entire network, NMS manages the equipments.

Each managed device includes Agent process, MIB and multiple managed objects that reside in the device. The NMS interacts with the Agent running on the managed device, and the Agent completes the instructions of the NMS through the operation of the MIB on the device end.

The network management model is shown in the following figure.



The following describes the main elements of the network management system:

- **NMS**

  NMS plays the role of administrator in the network. It is a system that adopts SNMP protocol to manage/monitor network devices and runs on the NMS server.

  - NMS can send a request to the Agent on the device to inquire or modify one or more specific parameter values.
  - NMS can receive the Trap information actively sent by the Agent on the device, so as to know the current state of the managed device.

- **Agent**

  Agent: Agent is an agent process in the managed device, which is used to maintain the information data of the managed device and respond to the request from the NMS, and report the administration data to the NMS that sending the request.

  - After receiving the request information from NMS, Agent completes the corresponding instructions through MIB table and responds the operation results to NMS.
  - In case of device failure or other events, the device will actively send information to NMS through Agent, and report the current state change of the device to NMS.

- **Management object**

  Management object refers to the managed object. Each device may contain multiple Management objects, which may be a piece of hardware (e.g: a interface board) in the device or a set of parameters configured on hardware or software (e.g: routing protocol) .

- **MIB**

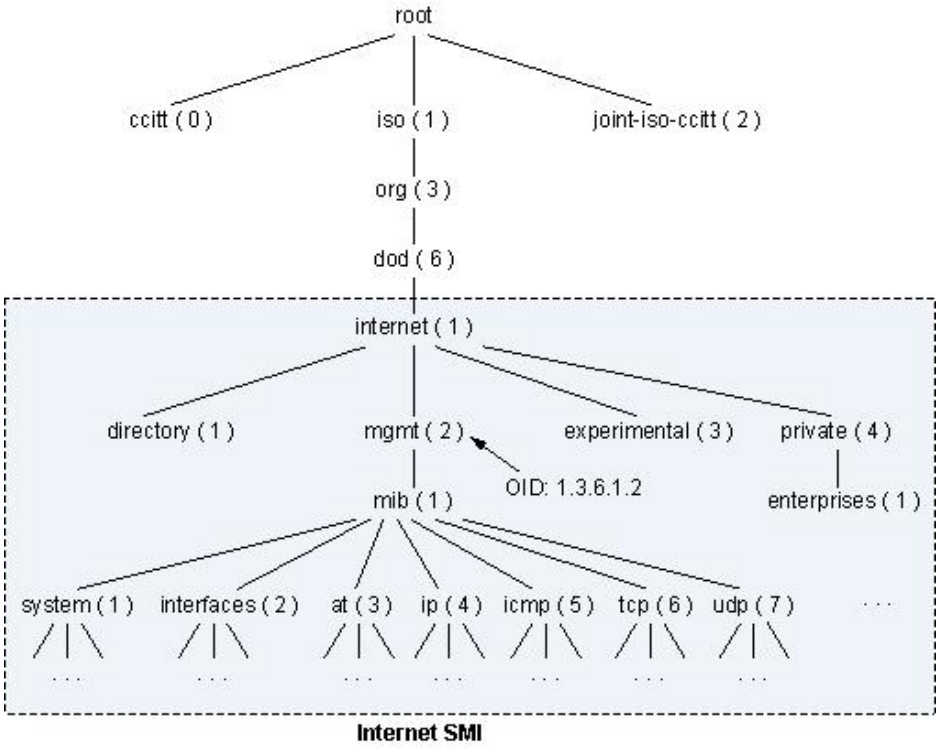
  MIB is a database, which indicates the variables maintained by managed devices (that is, information that can be queried and set by Agent). The MIB defines a series of attributes of the managed device in the database: the name of the object, the state of the object, the access rights of the object, and the data type of the object.

  With MIB, you can complete the following functions:

  - The Agent can get the current status information of the device by querying MIB.
  - Agent can set the state parameters of devices by modifying MIB.

The management information base of SNMP adopts a tree structure similar to DNS (Domain Name System), with its unnamed root at the top. The following figure shows a part of the management information base, which is also called the object naming tree. Every OID(object identifier) corresponds to a management object in the tree, for example, the OID of system is 1.3.6.1.2.1.1, and the OID of interfaces is 1.3.6.1.2.1.2. Through OID tree, the management information stored in it can be managed efficiently and conveniently, and it is also convenient to query the information in batches.

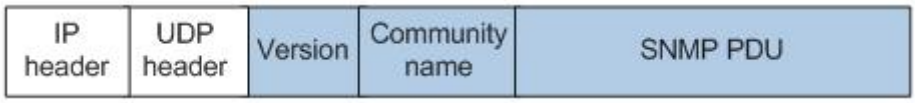
Especially, when the user is configuring Agent, MIB objects that NMS can access can be restricted through MIB view. The MIB view is actually a subset of MIBs.



23.2.2 SNMPv1/SNMPv2c

23.2.2.1 Message Structure of SNMPv1/SNMPv2c

As shown in the figure below, SNMPv1/SNMPv2c message is mainly composed of version, community name and SNMP PDU.



The main fields in the message are defined as follows:

- Version: indicates the version of SNMP. If it is SNMPv1 message, the corresponding field value is 0, and SNMPv2c is 1.
- Community name: used to complete authentication between Agent and NMS, in string form, which can be defined by user. Community names include "readable" and "writable". When executing GetRequest and GetNextRequest operations, "readable community name" is used for authentication. When executing the Set operation, the "writable community name" authentication is adopted.
- SNMPv1/SNMPv2c PDU: contains PDU type, request identifier, variable binding list and other information. There are several types of SNMPv1 PDU, including GetRequest PDU, GetNextRequest PDU, SetRequest PDU, Response PDU and Trap PDU. SNMPv2c PDU adds two types of GetBulkRequest PDU and InformRequest PDU based on SNMPv1.

To simplify, SNMP operations will be called Get, GetNext, Set, Response, Trap, GetBulk and Inform operations henceforth.

23.2.2.2 Operation Type of SNMPv1/SNMPv2c

As shown in the following table, SNMPv1/SNMPv2c specifies 7 types of operations to complete information exchange between NMS and Agent.

Operation	Description
Get	The Get operation can extract one or more parameter values from the Agent.
GetNext	The getNext operation extracts the value of the next parameter from the Agent in lexicographical order.
Set	The Set operation can set one or more parameter values of the Agent.

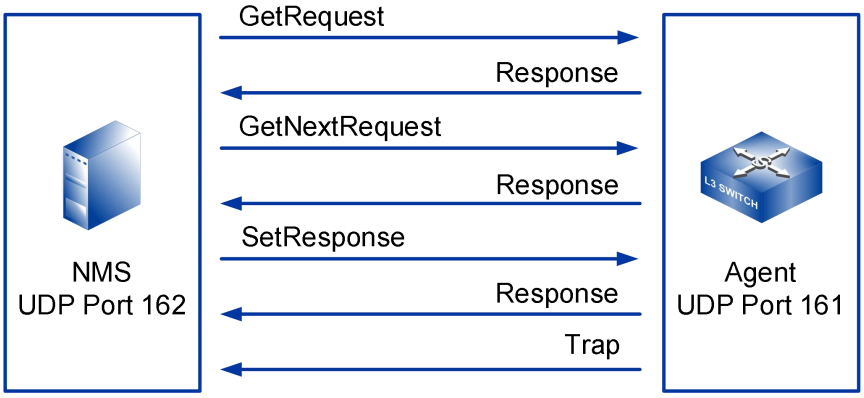
Operation	Description
Response	The Response operation can back to one or more parameter values. This operation is issued by the Agent, which is the response operation of GetRequest, GetNextRequest, SetRequest and GetBulkRequest. After receiving the Get/Set instruction from NMS, the Agent completes the corresponding query/modification operation through MIB, and then uses Response operation to respond the information to NMS.
Trap	Trap information is the information sent by the Agent to NMS to inform the management process of the situation on the device end.
GetBulk	GetBulk operation has achieved NMS's information group query toward managed devices.
Inform	InformRequest is also managed device sending warning to NMS proactively. Different from Trap alarm, NMS needs to reply InformResponse for confirmation after the managed device sends Inform warning.

Note

The SNMPv1 version does not support the GetBulk and Inform operations.

23.2.2.3 Working principle of SNMPv1/SNMPv2c

The working principles of SNMPv1 and SNMPv2c are basically the same. The working principle of SNMPv1/SNMPv2c is shown in the following figure.



- Get operation

  Assume that NMS wants to obtain the value of sysContact of MIB node of managed device, and uses the readable community name as public. The process is as follows:

  - NMS: send Get request message to Agent. Each field in the message is set as follows: the version number is the SNMP version used; Group name is public; PDU type in PDU is Get type, and binding variable is filled with MIB node name sysContact.
  - Agent: firstly, the version number and community name carried in the message are authenticated. After the authentication is successful, the Agent queries the sysContact node in MIB according to the request, obtains the value of sysContact, encapsulates it into PDU in the Response message, and sends a response to NMS; If the query is unsuccessful, Agent will send an error response to NMS.

- GetNext operation

  Assume that NMS wants to obtain the sysName value of the next node of the managed device MIB node sysContact, and uses the readable community name as public. The process is as follows:

  - NMS: send GetNext request message to Agent. Each field in the message is set as follows: the version number is the SNMP version used; Group name is public; PDU type in PDU is GetNext type, and binding variable is filled with MIB node name sysContact.
  - Agent: firstly, the version number and community name carried in the message are authenticated. After the authentication is successful, the Agent queries the next node sysName of sysContact in MIB according to the request, obtains the value of sysName, encapsulates it into PDU in Response message, and sends a response to NMS; If the query is unsuccessful, Agent will send an error response to NMS.

- Set Operation

  Assume that NMS wants to set the value of sysName of MIB node of managed device as Switch, and use the writable community name as private. The process is as follows:

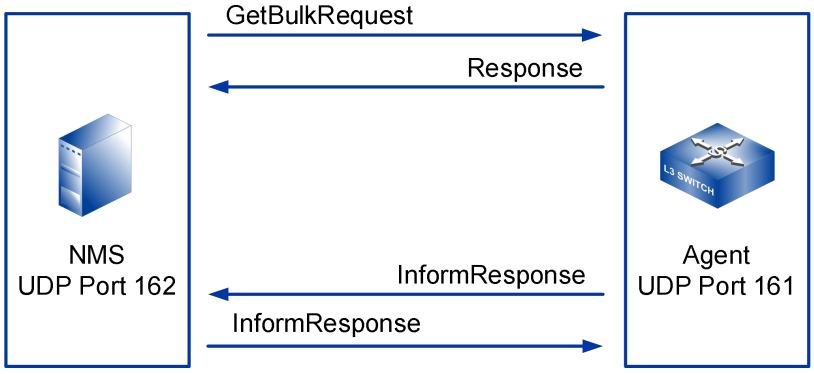
  - NMS: send Set request message to Agent. Each field in the message is set as follows: the version number is the SNMP version used; Group name is private; PDU type in PDU is Set type, and binding variable is filled with MIB node name sysContact and value Switch to be set.
  - Agent: firstly, the version number and community name carried in the message are authenticated. After the authentication is successful, the Agent sets the node corresponding to the management variable in the management information base MIB according to the request, and sends a response to the NMS after the setting is successful; If the setting is unsuccessful, the Agent will send an error response to NMS.

- Trap operation

  Trap does not belong to the basic operation of NMS on managed devices, it is the spontaneous behavior of managed devices. When the managed device reaches the trigger condition of alarm, it will send a Trap message to NMS through Agent to inform the abnormal situation on the device side, which is convenient for the network administrator to handle in time. For example, after the managed device is warm-started, the Agent will send the Trap of warmStart to NMS.

  This kind of Trap information is limited. The Agent will report to the management process only when the module on the device side reaches the predefined alarm triggering condition of the module. This method has the advantage of sending Trap information only when serious events occur, thus reducing the traffic generated by message interaction.

The new operation of SNMPv2c is shown in the following figure.



- GetBulk operation

  Implementation based on GetNext is equivalent to executing GetNext operations

several times continuously. On NMS, you can set the number of times that the managed device executes GetNext operation when interacting with a GetBulk message.

- Inform operation

  Inform operation is also a managed device that sends an active alert to the NMS. Different from trap alarm, NMS is required to confirm receipt after the managed device sends Inform alarm. If the managed device does not receive the confirmation message:

  - Save the alarm temporarily in the Inform cache.
  - The alarm is sent repeatedly until NMS confirms receipt of the alarm or the sending times reach the maximum retransmission times.
  - A corresponding alarm log is generated on the managed device.

  Therefore, using Inform alarm may occupy more system resources.

23.2.3 SNMPv3

23.2.3.1 Message Structure of SNMPv3

SNMPv3 defines a new message format, and its message structure is shown in the following figure.

The main fields in SNMP messages are defined as follows:

- Version: indicates the version of SNMP, and the corresponding field value of SNMPv3 message is 2.
- Header data: it mainly includes the description contents such as the maximum message size supported by the message sender and the security mode adopted by the message.
- Security parameters: including the relevant information of SNMP entity engine, user name, authentication parameters, encryption parameters and other security information.
- Context EngineID：SNMP unique identifier, which together with PDU type determines which application should be sent to.
- Context Name: used to determine MIB view of Context EngineID to managed devices.
- SNMPv3 PDU: It contains PDU type, request identifier, variable binding list and other information. In which SNMPv3 PDU include GetRequest PDU,

GetNextRequest PDU, SetRequest PDU, Response PDU, Trap PDU, GetBulkRequest PDU and InformRequest PDU.

23.2.3.2 SNMPv3 Architecture

SNMPv3 proposes a new SNMP architecture, which provides a general implementation model for various NMS based on SNMP, namely SNMPv3 entity. The SNMPv3 entity can be divided into SNMPv3 engine and SNMPv3 application, both of which are composed of several small modules.

The modular structure of SNMPv3 entity has the following advantages:

- Adaptability: It is suitable for various operating environments, which can manage the simplest network and meet the management requirements of complex networks.
- Convenient management: SNMP framework is composed of several subsystems or applications with relatively independent functions, so it can be managed conveniently. For example, if the system fails, the corresponding subsystem can be located according to the failure function type.
- Good extensibility: through SNMP entity, the system can be expanded conveniently. For example, in order to apply a new security protocol, a separate module can be defined in the security subsystem to support the protocol in SNMP.

With the adoption of USM(User-based Security Model) and VACM (view-based access control model), SNMPv3 has been improved its security.

- USM: provides authentication and data encryption services. To realize this function, NMS and Agent must share the same key.

  Authentication: authentication means that when an Agent or NMS receives information, it must first confirm whether the information comes from an authorized NMS or Agent and the information has not been changed during transmission. HMAC is defined in RFC2104, which is an effective tool for generating information verification codes by using secure hash functions and keys, and has been widely used in the Internet. There are two kinds of HMAC used by SNMP: HMAC-MD5-96 and HMAC-SHA-96. The hash function of the former is MD5 and takes 128-bit authKey as input. The hash function of the latter is SHA-1 and takes 160-bit authKey as input.

  Encryption: Encryption algorithm is mainly realized through symmetric key system, which uses the same key to encrypt and decrypt data. The encryption process is similar to authentication, and it also requires the management station and the agent to share the same key to encrypt and decrypt information. SNMP uses the following three encryption algorithms:

- DES: encrypt a 64-bit plaintext block with a 56-bit key.
- 3DES: use three 56-bit DES keys (total 168-bit keys) to encrypt plaintext.
- AES: use AES algorithm with key length of 128bit, 192bit or 256bit to encrypt plaintext.

Note

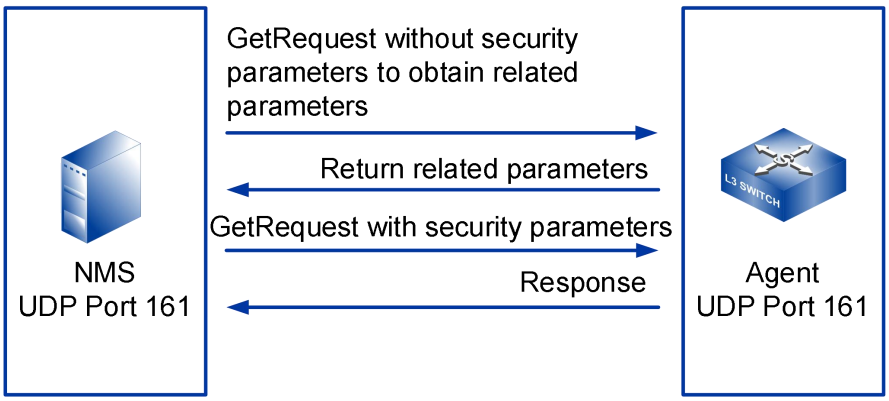
The security of these three encryption algorithms from high to low is AES, 3DES and DES. The encryption algorithms with high security have complex implementation mechanism and slow operation speed.

- VACM: Implement view-based access control for user groups or community names. Users must first configure a view and indicate permissions. Users can load this view to limit read and write operations, Inform or Trap when configuring user or user group or community name.

23.2.3.3 SNMPv3 Operating Principle

The realization principle of SNMPv3 is basically the same as SNMPv1/SNMPv2c, the only difference is that SNMPv3 adds authentication and encryption processing. The following describes the working principle of SNMPv3 by taking Get operation as an example.

Assume that NMS wants to obtain the value of sysContact of MIB node of managed devices, and uses authentication encryption, as shown in the following figure:



1 NMS: Send a Get request message without security parameters to the Agent, and obtain the Context EngineID, Context Name and security parameters (related information of SNMP entity engine) from the Agent.

2 Agent: responds to the request of NMS and feeds back the requested parameters to NMS.

3 NMS: Send the Get request message to Agent again. The settings of each field in

the message are as follows:

- Version: SNMPv3 version.
- Header data: indicates authentication and encryption.
- Security parameters: NMS calculates authentication parameters and encryption parameters through the configured algorithm. Fill these parameters and the acquired security parameters into the corresponding fields.
- PDU: fill the obtained Context EngineID and Context Name into the corresponding fields, set the PDU type to Get, fill the binding variable into the MIB node name sysContact, and encrypt the PDU with the configured encryption algorithm.

4 Agent: firstly, authenticate the message, and decrypt the PDU after passing the authentication. Upon successful decryption, the Agent queries the sysContact node in MIB according to the request, obtains the value of sysContact and encapsulates it into the PDU in the Response message, encrypts the PDU, and sends a response to NMS. If query is unsuccessful or authentication and decryption fail, Agent will send error response to NMS.

23.3 Configure SNMP

23.3.1 SNMP Enablement

【Command】

```
snmp-server enable traps
no snmp-server enable traps
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

snmp-server enable traps: command is used to enable SNMP server.

no snmp-server enable traps: command is used to disable SNMP server.

By default, SNMP function is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#snmp-server enable traps
Switch(config)#no snmp-server enable traps
```

23.3.2 SNMP View

【Command】

```
snmp-server view VIEWNAME OID ( included | excluded )
no snmp-server view VIEWNAME OID
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

VIEWNAME: the view name, with a value range of 1-32 bytes.

OID: OID MIB subtree of MIB object subtree, variable OID only allows digital input (such as 1.3.6.1).

Included: means that this MIB view includes the MIB subtree.

excluded: means that this MIB view excludes the MIB subtree.

【Description】

snmp-server view: command is used to create or update information about the MIB view to restrict the MIB objects that can be accessed by the NMS.

no snmp-server view: the command is used to cancel the current configurations.

A MIB is a collection of managed objects, and a MIB view is a subcollection of a MIB. User configuration can bind the group name/user name to the MIB view, thereby limiting the MIB objects that an NMS can access. The user can configure the MIB object to excluded or included within the view. Excluded means that the current view does not include all the modes of the MIB subtree; Included means that the current view includes all nodes of the MIB subtree.

By default, the view name is system. The OID included is 1.3.6.1.

SNMP community name or group name configuration needs to determine the MIB view permissions of the community name or group, related configuration can refer to the command snmp-server community, snmp-server group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#snmp-server view viewname 1.3.6.1.5 included
Switch(config)#no snmp-server view viewname 1.3.6.1.5
```

23.3.3 SNMP Community Name

【Command】

```
snmp-server community NAME {view VIEWNAME | } (ro | rw)
no snmp-server community {NAME | }
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

NAME: the team name, with a value range of 1-32 bytes.

View: MIB view name, this parameter is optional, if not entered, by default it is the default view.

Ro: read only means read-only access to MIB objects. Communities with read-only access can only view device information.

rw: read and write indicates read and write access to MIB objects, and communities with read and write permissions can configure devices.

【Description】

snmp-server community: command is used to set the community name, SNMP v1/v2c version uses the group name to restrict access rights. This command can be used to configure the group name, read or write view rights and access control policies. **no snmp-server community**: command is used to cancel group access name settings.

Normally, "public" is used as the name of the read permission group. For security reasons, it is recommended that network administrators configure other community names.

【Instance】

```
Switch> enable
Switch#configure terminal
```

```
Switch(config)#snmp-server community communityname view viewname
rw
Switch(config)#no snmp-server community communityname
```

23.3.4 SNMP Group

【Command】

```
snmp-server group NAME v3 (auth | noauth | priv ) [ (notify | read
| write ) VIEWNAME ]
no snmp-server group NAME v3 (auth | noauth | priv )
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

NAME: the group name, with a value range of 1-32 bytes.

v3: SNMP v3 version.

Auth: indicates that the message is authenticated but not encrypted.

Noauth: indicates that the message is neither authenticated nor encrypted.

Priv: indicates that the message is authenticated and encrypted.

VIEWNAME: view name, ranging from 1 to 32 bytes. By default, the Trap message view is not configured, meaning that the Agent does not send Traps to the NMS.

Read: specifies the read view of the group.

write: specifies the write and read view of the group.

VIEWNAME: view name

【Description】

snmp-server group: command is used to configure a new SNMP group and set the secure mode and corresponding SNMP view of the SNMP group.

no snmp-server group: command is used to delete a specified SNMP group. For SNMP v3, the group name and the security mode (authentication or not, encryption or not) together determine a group, with the same group name but different security mode are two different groups.

This system defaults to snmp v2, so there is no default configuration for group. If the view name for read is not specified in the command, it defaults to the default view.

【Instance】

```
Switch> enable
```

```
Switch#configure terminal
Switch(config)#snmp-server group groupname v3 priv read viewname
write viewname
Switch(config)#no snmp-server group groupname v3 priv
```

23.3.5 SNMP User

【Command】

```
snmp-server user USERNAME GROUPNAME [v3 [auth md5 MD5 [priv (aes
| des) PASSWORD] ] ]
no snmp-server user USERNAME GROUPNAME v3
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

USERNAME: user name, ranging from 1-32 bytes.

GROUPNAME: group name

v3: specifies that it is SNMP v3 version user, and defaults to v1 version user.

Auth: indicates that security mode requires authentication. If do not enter this parameter, the default is no authentication, no encryption mode.

md5: specifies the authentication protocol as the HMAC MD5 algorithm.

MD5: authentication password, string, value range of plaintext is 1 ~ 64 characters. If MD5 algorithm is adopted in ciphertext form, the authentication key is 32-bit hexadecimal number. If SHA algorithm is used, the authentication key is a 40-bit hexadecimal number.

priv: indicates that security mode requires authentication.

aes: the encryption algorithm is specified as AES (Advanced Encryption Standard), which has higher security than DES.

des: the encryption algorithm is specified as DES (Data Encryption Standard).

password: encrypted password, string, value range of plaintext is 1 ~ 64 characters. If MD5 algorithm is adopted in ciphertext form, the authentication key is 32-bit hexadecimal number. If SHA algorithm is used, the authentication key is a 40-bit hexadecimal number.

【Description】

snmp-server user: command is used to add a new user to an SNMP group.

no snmp-server user: command is used to delete a user of an SNMP group.

This command applies to SNMP v3 version. If the Agent interact with the message of NMS using SNMP v3 version, then SNMP v3 users need to be created. For the configured user to take effect, a group must be created first. Authentication and encryption are configured when the group is created, and the specific algorithm and password for authentication and encryption are configured when the user is created.

Notice

This command is used several times to configure the same user (that is, the user name is the same, no other parameters are required), and the configuration results are subject to the last configuration.

【Instance】

```
Switch> en
Switch#configure terminal
Switch(config)#snmp-server user admin groupname v3
Switch(config)#no snmp-server user username groupname v3
```

23.3.6 SNMP Trap Destination

【Command】

```
snmp-server host [ipv6] IP traps (version ( 1 | 2c )) NAME
no snmp-server host [ipv6] IP traps NAME
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

traps: specify the host to be Trap host.

IP: the IPv4 or IPv6 address of a host that accepts Traps.

1: represents SNMP v1 version.

2c: represents SNMP v2c version.

NAME: when there is a parameter version, NAME represents SNMPv1/v2c community. When there is no version parameter, NAME represents SNMPv3 user name.

【Description】

snmp-server host: command is used to set the destination host to receive SNMP Trap messages.

no snmp-server host: command is used to cancel the current configurations.

Depending on network management needs, users can configure multiple destination hosts to receive Trap messages through this command.

If a device is needed to send Trap messages, the snmp -server host command should be used in conjunction with the snmp -server enable trap command (the default is to send all traps).

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#snmp-server host 192.168.5.123 traps version 2c communityname
Switch(config)#no snmp-server host 192.168.5.123 traps name
```

23.3.7 View information about SNMP

【Command】

show snmp [community | (group [WORD]) | host | sub-agent | (user [WORD]) | view]

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

WORD: the specified group name or user name.

【Description】

show snmp: view SNMP status information.

show snmp community: view SNMP community information.

show snmp group [WORD]: view SNMP group or specified SNMP group information.

show snmp host: check the information of the destination host receiving the Trap message.

show snmp sub-Agent: view SNMP sub-agent information.

show snmp user [WORD]: view SNMP user or specified user information.

show snmp view: view SNMP MIB view information.

【Instance】

```
Switch#show snmp view
View-name                      Oid                    Type
system                         1.3.6.1                included
```

24 LLDP Configuration

24.1 Overview

LLDP (Link Layer Discovery Protocol) is a link layer discovery protocol defined in IEEE 802.1ab. LLDP is a standard layer-2 discovery method, which can organize the management address, device identification, interface identification and other information of local devices and publish it to its neighbor devices. After receiving the information, the neighbor devices save it in the form of standard MIB(Management Information Base) for the network management system to query and judge the communication status of links.

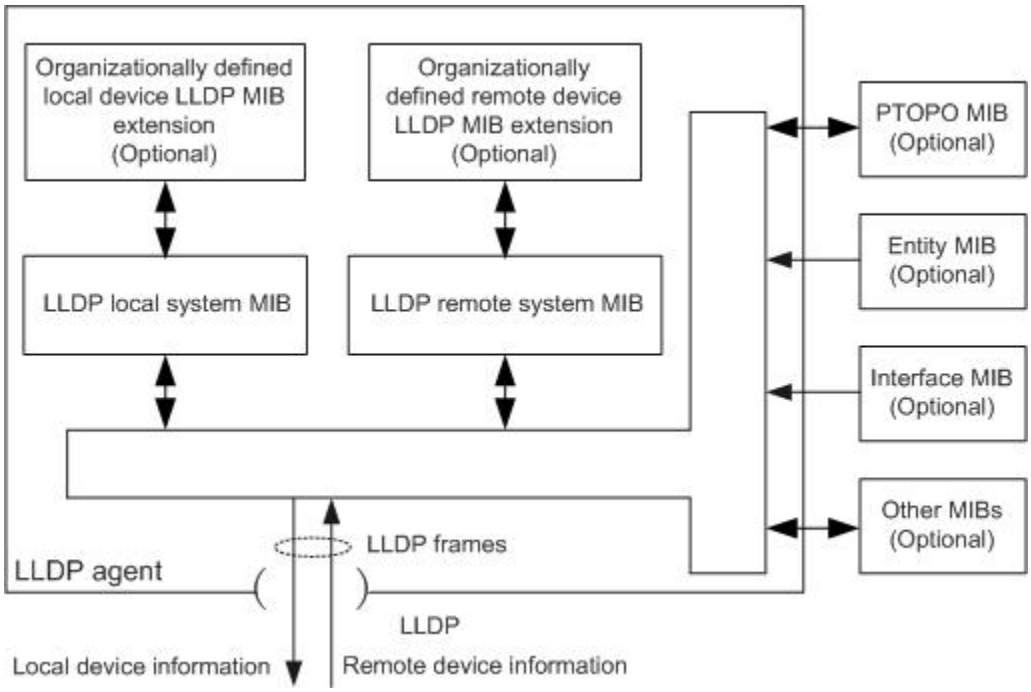
With the increasing scale of the network, there are many kinds of network devices, and their configurations are complicated, so the requirements for network management ability are getting higher and higher. Most traditional network management systems can only analyze the Layer 3 network topology, and cannot determine the detailed topology information of network device and whether there is configuration conflict. Therefore, it is necessary to have a standard Layer 2 information exchange protocol.

LLDP provides a standard link layer discovery method. The device layer 2 information obtained by LLDP can quickly obtain the topological state of connected device; Display the paths among clients, switches, routers, application servers and network servers. Detect the configuration conflicts between devices and query the reasons of network failure. Enterprise network users can use the network management system to monitor the link state of the device supporting LLDP protocol, and quickly locate the fault when the network fails.

24.2 Principles

24.2.1 Working Principle

LLDP can organize the information of local devices and publish it to its own remote devices, and the local devices save the received remote device information in the form of standard MIB. The working principle is shown in the following figure.



The basic realization principle of LLDP is:

1 The LLDP module updates its own LLDP local system MIB and the LLDP extension MIB customized by the local device by interacting LLDP agent with the physical topology MIB, entity MIB, interface MIB and other types of MIB on the device.

2 Encapsulate the local device information into LLDP frame and send it to the remote device.

3 Receive LLDP frame sent by remote devices, update own LLDP remote system MIB and LLDP extended MIB customized by remote devices.

4 By sending and receiving LLDP frames through the LLDP proxy, the device can clearly know the information of the remote device, including which interface of the remote device is connected and the MAC address of the remote device.

The LLDP local system MIB is used to store local device information. Include device ID, interface ID, system name, system description, interface description, network management address and other information.

The LLDP remote system MIB is used to store remote device information. Include device ID, interface ID, system name, system description, interface description, network management address and other information.

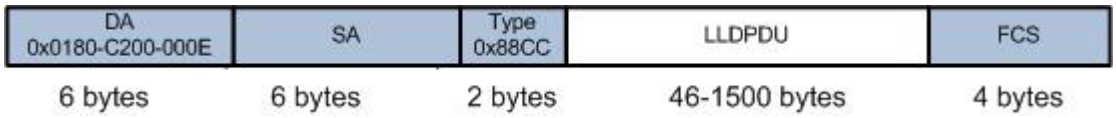
The LLDP agent performs the following tasks:

- Maintain LLDP local system MIB and LLDP remote system MIB.
- When the local state changes, the MIB information of LLDP local system is extracted and sent to the remote device. Under the condition that the status information of the local device does not change, the MIB information of the LLDP local system is extracted according to a certain period and sent to the remote device.
- Identify and process the received LLDP frame.
- Send LLDP alarm to the network administrator when the state of LLDP local system MIB or LLDP remote system MIB changes.

24.2.2 Message Structure

Ethernet message encapsulated with LLDP (LLDP Data Unit) is called LLDP message.

The LLDP message structure is shown in the following figure.



The meaning of each field is as follows:

- DA(Destination MAC Address): destination MAC address, which is a fixed multicast MAC address 0x0180-C200-000E.
- SA(Source MAC Address): source MAC address, which is the MAC address of the sender.
- Type: message type. the value of this field in LLDP message is 0x88CC.
- LLDPDU: LLDP data unit, the main body of LLDP information exchange.
- FCS: frame check sequence.

24.2.2.1 LLDPDU

LLDPDU is the data unit of local information encapsulated in LLDP message. Before forming LLDPDU, local information is encapsulated into TLV(Type/Length/Value) format, and then several TLVs are combined into one LLDPDU, which is

encapsulated in the data part of LLDP message for transmission. The LLDPDU structure is shown in the following figure.

As shown in the above figure, Chassis ID TLV, Port ID TLV, Time to Live TLV and End of LLDPDU TLV are required TLVs. The rest are optional TLVs, which can be defined by the device whether they are included in the LLDPDU.

When the state of the interface changes (LLDP and interface shutdown is enabled), the interface will send an LLDP message to the neighbor device, in which the Value of the Time To Live TLV field is 0, and this message is called shutdown message. Shutdown message does not contain any optional TLV.

24.2.2.2 TLV Structure

A TLV is a unit that constitutes an LLDPDU, and each TLV represents a piece of information.

The structure of TLV is shown in the following figure.

The meaning of each field is as follows:

- TLV Type: the type of TLV. The type value of each TLV is different, such as the type value of End of LLDPDU TLV is 0, and the type value of Chassis ID TLV is 1.
- TLV Length: the length of TLV, accounting for 9 bits.
- TLV Value: the value of the TLV, the first byte refers to the subtype of the TLV, and the remaining bytes are the real values of the TLV.

24.2.2.3 TLV type

The types of TLV that LLDP can encapsulate include basic TLV, TLV defined by 802.1 organizations, TLV defined by 802.3 organizations, and MED (Media Endpoint Discovery) TLV. The basic TLV is a group of basic TLV for managing device, while the TLV defined by 802.1 organization, the TLV defined by 802.3 organization and the MED TLV are TLV defined by standard organization or other institutions, which are used to enhance the management function of device, and can be sent in LLDPDU according to actual needs.

- Basic TLV

  In the basic TLV, there are four types of TLV which are necessary for realizing LLDP function, that is, they must be published in LLDPDU.

TLV name	Note	Is it necessary to publish
Chassis ID TLV	The bridge MAC address of the sending device.	Yes
Port ID TLV	Identifies the port of the sending end of LLDPDU, and the content is the port name.	Yes
Time To Live TLV	The lifetime of this device information on neighboring devices.	Yes
End of LLDPDU TLV	Marks the end of LLDPDU.	Yes
Port Description TLV	Description string of the Ethernet port.	No
System Name TLV	Display the device name.	No
System Description TLV	System description.	No
System Capabilities TLV	Main functions of the system and which main functions are enabled.	No
Management Address TLV	Address for the network management system to identify and manage network device. Management address can definitely mark a device, which is beneficial to the drawing of network topology and network management.	No

- TLV defined by IEEE 802.1 organization

TLV name	Note
Port VLAN ID TLV	Port VLAN ID.
Port And Protocol VLAN ID TLV	Protocol VLAN ID of the port.
VLAN Name TLV	VLAN name of that port.

TLV name	Note
Protocol Identity TLV	Protocol types supported by the port.

- TLV defined by IEEE 802.3 organization

TLV name	Note
EEE TLV	Whether the port supports EEE(Energy Efficient Ethernet) function.
Link Aggregation TLV	Whether the port supports link aggregation and whether link aggregation is enabled.
MAC/PHY Configuration/Status TLV	The port's rate and duplex status, whether port rate auto-negotiation is supported, whether auto-negotiation is enabled, and the current rate and duplex status.
Maximum Frame Size TLV	The maximum frame length supported by the port is the MTU (Max Transmission Unit) of the port.
Power Via MDI TLV	The power supply capability of the port, such as whether PoE is supported, it is a power supply device or a power receiving device.

- MED TLV

  MED TLV provides many advanced applications for VoIP(Voice over IP), including basic configuration, network policy configuration, address information, directory management, etc., which meets the requirements of different manufacturers of voice equipment in terms of cost-effectiveness, easy deployment and easy management, and solves the problem of deploying voice equipment in Ethernet, providing convenience for producers, sellers and users of voice equipment.

TLV name	Note
LLDP-MED Capabilities TLV	Device type of current device and LLDP-MED TLV type that can be encapsulated in LLDPDU.

TLV name	Note
Inventory TLV	The manufacturer of the device.
Location Identification TLV	Location identification information for other devices to discover the location of the device.
Network Policy TLV	The VLAN ID, layer 2 priority and DSCP value of Voice VLAN.
Extended Power-via-MDI TLV	Power supply capability of current device.
Hardware Revision TLV	Hardware version of ME(Media Endpoint) device.
Firmware Revision TLV	Hardware version of ME device.
Software Revision TLV	Software version of ME device.
Serial Number TLV	Serial number of ME device.
Model Name TLV	The Model Name of the ME device.
Asset ID TLV	The asset identifier of the ME device.

24.2.3 Message Transmission Mechanism

LLDP message sending mechanism

When the LLDP function is enabled, the device will periodically send LLDP messages to neighboring devices. If the local configuration of the device changes, the LLDP message is sent immediately to inform the neighbor device of the change of local information as soon as possible. For preventing abounding LLDP sending caused by frequent changes of local information, next message should be delayed to send out after sending a LLDP message.

LLDP message receiving mechanism

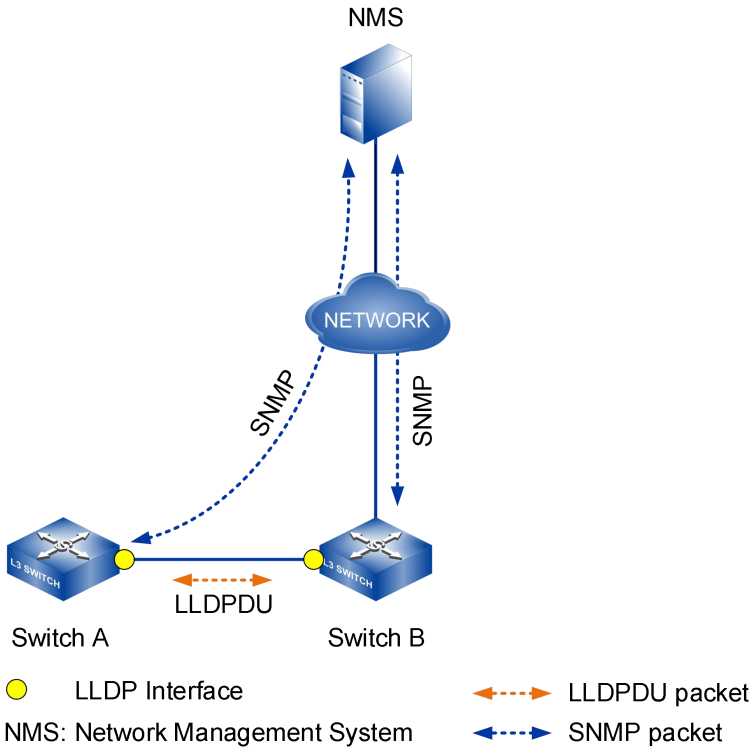
When enabling LLDP function, the device will check the validity of the received LLDP message and the TLV carried by it. After checking, the neighbor information will be saved to the local device, and the aging time of neighbor information in the local device will be set according to the TTL value carried by TLV in the LLDPDU message. If the TTL value in the received LLDPDU is equal to zero, the neighbor information will be aged immediately.

24.2.4 Networking Mode

Several common networking modes in LLDP applications;

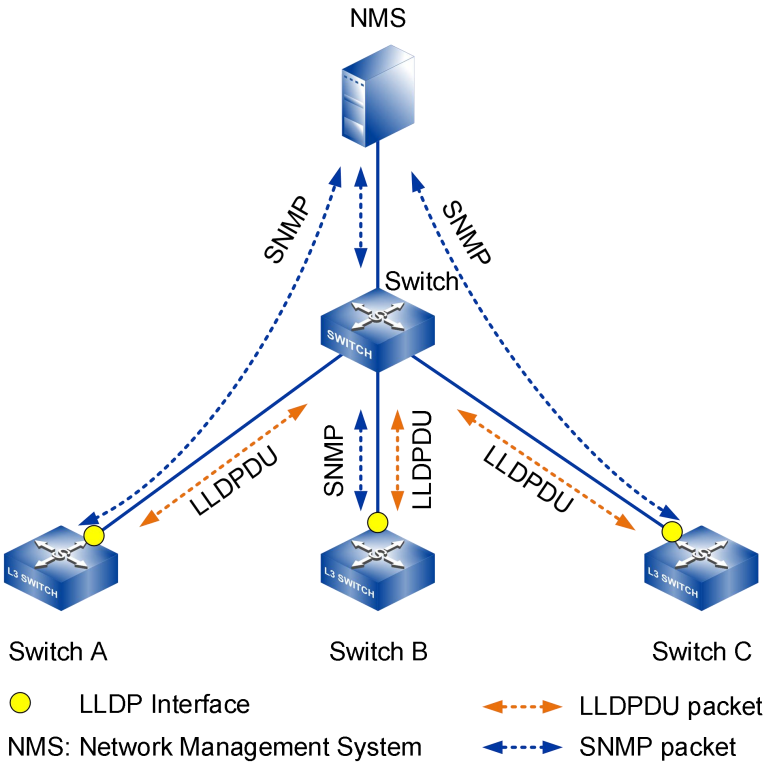
- Single neighbor networking mode

  Single-neighbor networking mode refers to the situation that the interfaces of switch devices are directly connected without crossing any devices in the middle, and the interface has only one neighbor device. Single neighbor networking is shown in the following figure. SwitchA and SwitchB are directly connected, and each interface of SwitchA and SwitchB has only one neighbor.



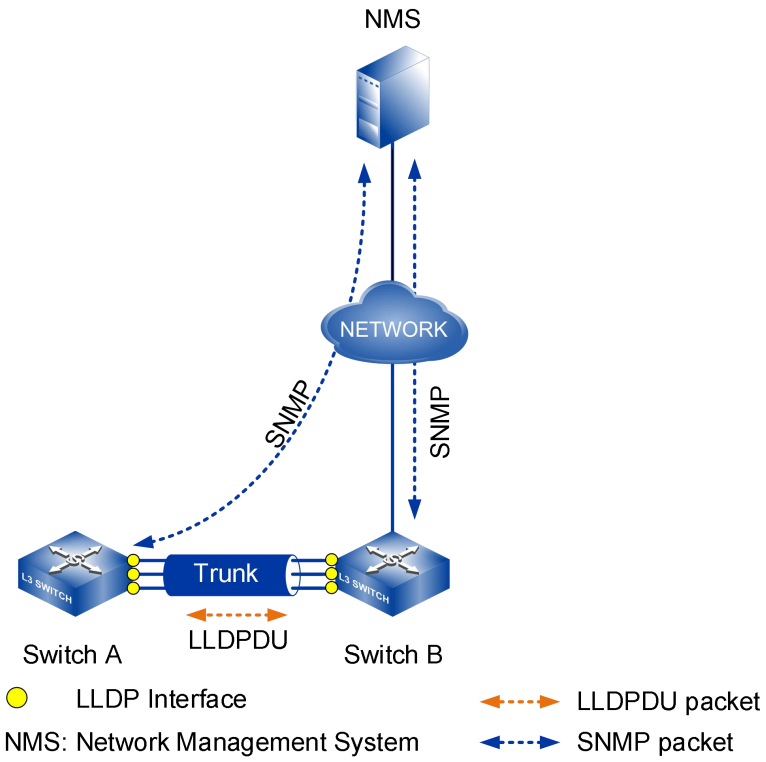
- Multi-neighbor networking mode

  Multi-neighbor networking mode means that the interfaces of switch equipment are not directly connected, and then each interface has more than one neighbor. The multi-neighbor networking is shown in the following figure, and SwitchA, SwitchB and SwitchC are connected by Switch (Switch needs to support LLDP message transparent transmission). In this way, the interfaces of SwitchA, SwitchB and SwitchC all have more than one neighbor.



- Link aggregation networking mode

  Link aggregation networking mode means that there is link aggregation between the interfaces of switch equipment, and the interfaces are directly connected, and each interface between link aggregation has only one neighbor device. As shown in the figure below, there is link aggregation between SwitchA and SwitchB, and each interface of SwitchA and SwitchB has only one neighbor.



24.3 Configure LLDP

24.3.1 LLDP Enablement

【Command】

lldp enable

no lldp enable

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

lldp enable: command is used to enable the LLDP function.

no lldp enable: the command is used to turn off lldp function.

By default, globle LLDP function is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#lldp enable
```

24.3.2 LLDP Port Operating Mode

【Command】

```
lldp admin-status (tx-enable | rx-enable | txrx-enable | disable )
no lldp admin-status disable
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

tx-enable: working mode is Tx, only sending and not receiving LLDP message.

rx-enable: work mode is Rx, it only receives LLDP message and not transmit it.

txrx-enable: work mode is TxRx, it transmits LLDP message as well as receive it.

Disable: work mode is Disable, it neither transmits nor receives LLDP message.

【Description】

lldp admin-status: command is used to configure the lldp working mode of the port.

no lldp admin-status disable: command is used to restore the default working mode of the port.

By default, the working mode of LLDP works in TxRx when global LLDP is enabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#lldp admin-status tx-enable
```

24.3.3 Time Interval of Sending LLDP Message

【Command】

```
lldp timer tx-interval <INTERVAL>
no lldp timer
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

INTERVAL: the time interval between ports to send LLDP message, ranging from 5-300 in seconds.

【Description】

lldp timer tx-interval: command is used to set the time interval for sending LLDP message.

no lldp timer: command is used to restore the default packet time interval for LLDP.

By default, the interval between LLDP message is 30 seconds

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#lldp timer tx-interval 50
```

24.3.4 LLDP Interface Management Address

【Command】

lldp management-address A.B.C.D

no lldp management-address

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: administrative address published in LLDP message.

【Description】

lldp management-address: command is used to configure the management address published in the LLDP message.

no lldp management-address: command is used to restore the default management address published in the LLDP message.

The management address released by the port in the LLDP message defaults to the main IP address of the smallest VLAN of the VLANs this port is in. If the VLAN is not configured with a main IP address, it will be 0.0.0.0.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#lldp management-address 2.2.2.2
```

24.3.5 Encapsulation Format of LLDP Message

【Command】

```
lldp frame-format (snap | ethernet2)
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

snap: encapsulation format of LLDP message is snap.

ethernet2: the encapsulation format of LLDP message is Ethernet2

【Description】

lldp frame-format: command is used to configure the encapsulation format of LLDP message.

By default, the encapsulation format of LLDP message is Ethernet2.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#lldp frame-format snap
```

24.3.6 Display LLDP Neighbor Information

【Command】

```
show lldp neighbor-information (brief | )
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

Brief:displays a summary of the neighbor device, or neighbor information of all ports without this parameter.

【Description】

show lldp neighbor-information: command is used to display information about the neighbor device.

【Instance】

```
Switch> enable
Switch#show lldp neighbor-information
LLDP neighbor information of port ge2
----------------------------------------
Neighbor index                      : 1
Update time                         : 1hours 57minutes 40seconds
Ageing time                         : 114seconds
Chassis ID type                     : MAC Address
Chassis ID                          : 0022.6f55.5556
Port ID type                        : Interface Name
Port ID                           : ge10
Time to live                        : 120 seconds
Port description                    : ge10
System name                       : SW5
System capabilities supported             : Bridge/Switch,Router
System capabilities enabled               : Bridge/Switch,Router
Management address subtype          : IPv4
Management address                 : 192.168.1.254
Interface number subtype            : System Port Number
Interface number                    : 5010
Object ID                           : Standard LLDP MIB
MAC/PHY Configuration/Status        :
   Auto-Negotiation supported       : Yes
   Auto-Negotiation enabled         : Yes
   Operational MAU type             : 1000BASE-T full duplex mode
Link Aggregation                    :
   Link aggregation supported       : Yes
   Link aggregation enabled         : No
```

```
   Aggregated port ID                : 0
Maximum Frame Size                   : 1518
Port VLAN ID                         : 1
---------------------------------------

LLDP Neighbors Number                : 1

Switch#show lldp neighbor-information brief
Local  Intf       Neighbor  System  Name       Neighbor  Port  ID
Ageing-time(s)
ge2            SW5                 ge10            91

LLDP Neighbors Number             : 1
```

24.3.7 Display LLDP Statistics Information

【Command】

show lldp statistics (interface IFNAME |)

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

interface IFNAME: displays statistics information of the specified port.

【Description】

show lldp statistics: command is used to display statistics for all ports, or for the specified port.

【Instance】

```
Switch> enable
Switch#show lldp statistics
Global LLDP traffic statistics:
       Total frames out: 268
       Total ages out: 0
       Total frames discarded: 0
       Total frames received in error: 0
       Total frames received in: 260
       Total frames TLVs discarded: 0
       Total frames TLVs unrecognized: 0
```

```
Switch#show lldp statistics ge2
Interface ge2 LLDP traffic statistics:
       Total frames out: 269
       Total ages out: 0
       Total frames discarded: 0
       Total frames received in error: 0
       Total frames received in: 260
       Total frames TLVs discarded: 0
       Total frames TLVs unrecognized: 0
```

24.3.8 Display LLDP Local Information

【Command】

show lldp local-information (interface IFNAME |)

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

Interface IFNAME: displays local information of the specified port.

【Description】

show lldp local-information: command is used to display all LLDP local information, or LLDP local information of the specified port.

【Instance】

```
Switch> enable
Switch#show lldp local-information
LLDP local-information of port ge2:
       Chassis ID subtype : MAC address
       Chassis ID        : 0022.6f01.cca3
       Port ID subtype   : Interface name
       Port ID          : ge2
       Port description  : ge2

       Management address type          : IPv4
       Management address              : 192.168.1.254
       Management address interface type : ifIndex
       Management address interface ID   : 5002
```

```
Management address OID          : 0

Port VLAN ID(PVID) : 1

Port and protocol VLAN ID(PPVID) : 0
Port and protocol VLAN supported : not supported
Port and protocol VLAN enabled   : no enabled

VLAN name of VLAN 1 : default

Link aggregation supported : supported
Link aggregation enabled   : not enabled
Aggregated port ID         : 0

Auto-negotiation supported     : supported
Auto-negotiation enabled       : enabled
PMD auto-negotiation advertised :
   10BASE-T half duplex mode
   10BASE-T full duplex mode
   100BASE-TX half duplex mode
   100BASE-TX full duplex mode
   1000BASE-T half duplex mode
   1000BASE-T full duplex mode
Operational MAU type          : speed(1000)/duplex(full)
```

24.3.9 Display LLDP Status Information

【Command】

show lldp status (interface IFNAME |)

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

Interface IFNAME: displays state information of the specified port.

【Description】

show lldp status: command is used to display global LLDP status information, or LLDP status information on the specified port.

【Instance】

```
Switch> enable
Switch#show lldp status
LLDP running-information
       System running status          : Running
       System description             : Switch
       Transmit interval              : 30 s
       Hold multiplier                : 4
       Reinit delay                   : 2 s
       Transmit delay                 : 2 s
       Notification enable            : Enable
       Notification Interval          : 5 s

Switch#show lldp status interface ge2
Interface[ge2] lldp status
       Port status of LLDP            :  Enable
       Admin status                   :  Rx_Tx
       Trap flag                      :  No
       Number of neighbors            :  1
       Number of sent optional TLV    :  9
```

25 QOS Configuration

25.1 Overview

25.1.1 QoS Introduction

QoS is used to evaluate the ability of service providers to meet customer service needs. By configuring QoS, the network traffic of enterprises can be regulated, network congestion can be avoided and managed, and the loss rate of messages can be reduced. At the same time, it can also provide dedicated bandwidth for enterprise users or provide differential services for different services (voice, video, data, etc.).

QoS Background

The popularity of the network and the diversification of services make the Internet traffic surge, resulting in network congestion, increasing forwarding delay, and even packet loss in severe cases, resulting in the decline of service quality or even unavailability. Therefore, to carry out these real-time services on the network, we must solve the problem of network congestion. The best way to solve the network congestion is to increase the network bandwidth, but considering the cost of operation and maintenance, this is unrealistic. The most effective solution is to apply a "guaranteed" strategy to manage the network traffic.

Under this background, QoS technology developed. QoS (Quality of Service) provides end-to-end service quality guarantee based on the requirements of different services. QoS is a tool to effectively utilize network resources, which allows different traffic to compete for network resources unequally, and voice, video and important data applications can be served preferentially in network devices. QoS technology is applied more and more in today's Internet, and its role is becoming more and more important.

25.1.2 Priority Mapping

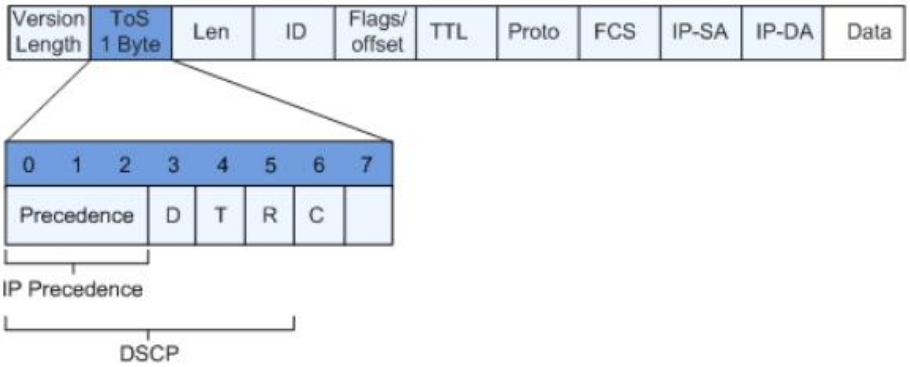
Priority mapping is used to realize the conversion between the QoS priority carried by the message and the internal priority of the device (also called local priority, which is the priority for distinguishing the message service level within the device), so that the device can provide differentiated QoS service quality according to the internal priority. Users can use different QoS priority fields in different networks according to network planning, such as EXP in MPLS network, 802.1p in VLAN network and DSCP in IP network. When a message passes through different networks, in order to keep the priority of the message, it is necessary to configure the mapping relationship of these priority fields on the devices connected to different networks. When the device is connected to different networks, the external priority fields (including MPLS EXP, 802.1p and DSCP) of all messages entering the device are mapped to the internal priority. When a device sends a message, it maps the internal priority to an external priority field.

QoS priority field

In order to provide different QoS service quality for different services on the Internet, people record QoS information according to some fields in the message header, so that each device in the network can provide different QoS according to this information. These QoS-related message fields include:

- Precedence field

As defined in RFC 791, the 8-bit Type of Service (ToS) field in an IP packet header contains a 3-bit IP precedence field. The following figure shows the Precedence field in an IP packet.



Bits 0 ~ 2 represent the Precedence field, which represents eight priorities of message transmission. The values are 7, 6, 5, 4, 3, 2, 1 and 0 in order of priority from high to low. High priorities are 7 and 6, which are often reserved for routing or updating network control communication. User-level applications can only use 0 ~ 5.

In addition to the Predecence field, the ToS field also includes three bits: D, T and R:

- D bit represents Delay requirement (0 represents normal delay and 1 represents low delay).
- T bit represents Throughput (0 represents normal throughput and 1 represents high throughput).
- R bit represents Reliability (0 represents normal reliability and 1 represents high reliability).

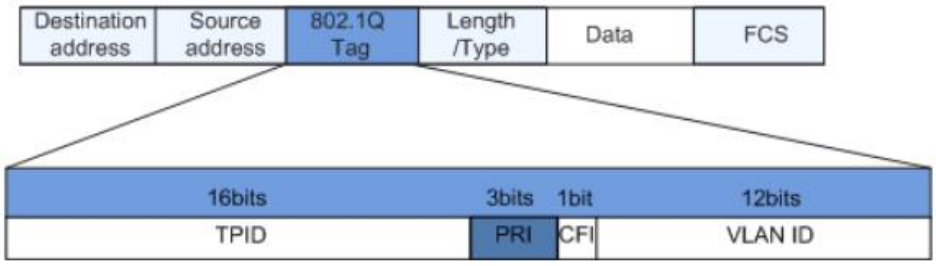
- DSCP field

RFC1349 redefines the ToS field in IP message, and adds C bit to indicate Monetary Cost. Later, IETF DiffServ Working Group redefined bits 0 ~ 5 in ToS field of IPv4 header as DSCP in RFC2474, and renamed ToS field as DS byte. The position of the DSCP in the packet is shown in the figure above.

The first 6 bits (0 ~ 5 bits) of DS field are used as DSCP(DS Code Point), and the last 2 bits (6 bits and 7 bits) are reserved bits. The first 3 bits (0 ~ 2 bits) of DS field are Class Selector Code Points (CSCP), and the same CSCP value represents Class I DSCP. A DS node selects the corresponding PHB according to the DSCP value.

- The 802.1p priority in VLAN header

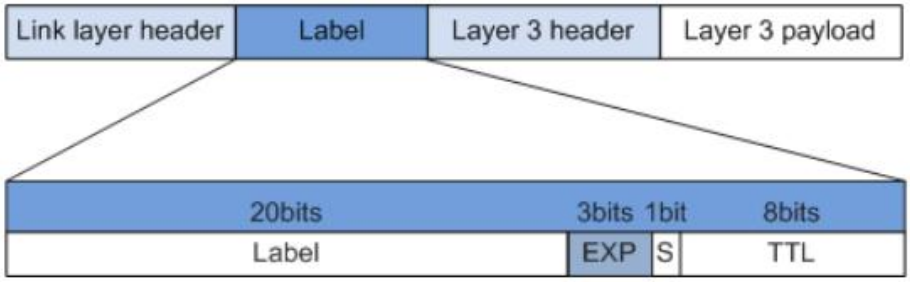
Generally, VLAN frames are exchanged between layer 2 devices. According to IEEE 802.1Q definition, the PRI field (i.e., 802.1p priority) or CoS field in VLAN frame header identifies the quality of service requirements. PRI field locations in VLAN frames are shown below.



The 802.1Q header contains a PRI field with a length of 3 bits. The PRI field defines eight service priority CoS, which are 7, 6, 5, 4, 3, 2, 1 and 0 in order of priority from high to low.

- MPLS EXP field

Compared with ordinary IP messages, MPLS messages add label information. The length of the label is 4 bytes, and the package structure is shown in the following figure.



There are 4 fields for labels:

- Label: 20 bits, label value field, pointer for forwarding.
- Exp: 3 bits, reserved field, used for extension, now commonly used as CoS.
- S: 1 bit, stack bottom identification. MPLS supports the hierarchical structure of labels, that is, multiple labels. When S value is 1, it indicates that it is the lowest label.
- TTl: 8 bits, which has the same meaning as TTL(Time To Live) in IP packet.

For MPLS messages, the EXP domain in label information is usually regarded as the CoS domain of MPLS messages, which is equivalent to the ToS domain of IP networks, and is used to distinguish the service level of data traffic to support DiffServ of MPLS networks. EXP field indicates 8 transmission priorities, and the values are 7, 6, ..., 1 and 0 in order of priority from high to low.

- In an IP network, the service level is identified by the IP priority or DSCP of the IP message. However, for MPLS networks, because the IP header of the packet is invisible to LSR(Label Switching Router) devices, it is necessary to mark the EXP field of MPLS packets at the edge of MPLS networks.
- By default, at the edge of MPLS network, the IP priority of IP message is directly copied to EXP domain of MPLS message. However, in some cases, if the ISP does not trust the user network, or if the differential service category defined by the ISP is different from the user network, the EXP domain of the MPLS message can be reset according to the internal service level and a certain classification strategy, while the ToS domain of the IP message remains unchanged during the forwarding process of the MPLS network.
- In the intermediate nodes of MPLS network, the packets are classified according to EXP domain, and congestion management, traffic supervision or traffic shaping are realized.

25.1.3 Flow Monitoring, Traffic Shaping and Interface Speed Limit

Traffic monitoring, traffic shaping and interface speed limit can limit traffic and improve the efficiency of network resources by monitoring the rate of traffic entering the network, thus ensuring better service for users.

When the sending rate of messages is higher than the receiving rate, or the interface rate of downstream device is lower than that of upstream device, it may cause network congestion. If the traffic sent by users is not limited, a large number of users' burst traffic data will make the network more crowded. In order to make the limited network resources serve users more effectively, it is necessary to limit the traffic flow of users.

Traffic supervision, traffic shaping and interface speed limit are flow control strategies that restrict traffic and resource usage by monitoring traffic specifications.

Traffic Policing

TP(Traffic Policing) can monitor the rate of different traffic entering the network, punish the excess traffic, and limit the incoming traffic within a reasonable range, thus protecting the network resources and the interests of users.

Traffic Shaping

TS(Traffic Shaping) is a measure to actively adjust the output rate of traffic. Traffic shaping cuts the irregular upstream flow and fills the valley, so that the flow output is relatively stable, thus solving the congestion problem of downstream equipment.

Line Rate

LR(Line Rate) can limit the total rate of sending or receiving all messages on an interface. The interface speed limit function can simplify the configuration when it is not necessary to distinguish message types but to limit all traffic rates through the interface.

25.1.4 Congestion Avoidance and Congestion Management

Congestion avoidance relieves network overload by specifying message dropping strategy, and congestion management ensures that high-priority services are processed first by specifying message scheduling order.

The quality of service problem faced by traditional networks is mainly caused by congestion. Congestion refers to a phenomenon that the rate drops and extra

extension is introduced due to insufficient network resources. Congestion will cause delay in message transmission, low throughput rate and large consumption of resources. In the complex environment where IP packet switching and multiple services coexist, congestion is very common.

Congestion avoidance and congestion management are two flow control methods to solve network congestion.

Congestion Avoidance

Congestion avoidance refers to a flow control mechanism that monitors the usage of network resources (such as queues or memory buffers), actively discards messages when congestion occurs or tends to increase, and relieves network overload by adjusting network flow.

The device supports the following congestion avoidance functions:

- Tail discard

The traditional discarding strategy adopts the method of tail discarding, which treats all messages equally and does not distinguish the service level of messages. When congestion occurs, the data messages at the end of the queue will be discarded until the congestion is relieved.

This dropping strategy will cause TCP global synchronization. The so-called TCP global synchronization phenomenon means that when multiple queues discard multiple TCP connection messages at the same time, some TCP connections will enter congestion avoidance and slow start state at the same time, reducing traffic to relieve congestion; Then these TCP connections will have a traffic peak at the same time. Repeatedly, the network traffic increases and decreases, which affects the link utilization.

By default, the interface adopts the discard policy of tail discard.

- WRED

WRED (Weighted Random Early Detection) randomly discards messages based on discarding parameters. Considering the benefits of high-priority messages and the relatively small probability of being discarded, WRED can specify different discarding strategies for messages of different services. In addition, by randomly discarding messages, multiple TCP connections can reduce the sending speed at different times, thus avoiding the phenomenon of TCP global synchronization.

WRED technology sets the upper and lower thresholds for the length of each queue, and stipulates that:

  - When the queue length is less than the lower threshold, the message is not discarded.

- When the queue length is greater than the upper threshold, all newly received messages are discarded.
- When the queue length is between the lower threshold and the upper threshold, the newly received message is discarded randomly. The method is to assign a random number to each newly received message, and compare the random number with the discard probability of the current queue. If it is less than the discard probability, the message is discarded The longer the queue, the higher the probability of message being discarded.

Congestion Management

Congestion management refers to a flow control mechanism that adjusts the scheduling order of messages to meet the high QoS service of delay-sensitive services when network congestion occurs intermittently and delay-sensitive services require higher QoS service than other services.

The device supports the following congestion management functions:

- PQ scheduling

PQ(Priority Queuing) scheduling means scheduling in strict accordance with the order of queue priorities. Only when all the messages in the high priority queue are scheduled, the low priority queue will have a scheduling opportunity.

PQ scheduling method is adopted to put delay-sensitive services into high priority queues and other services into low priority queues, thus ensuring that delay-sensitive services are scheduled preferentially.

The disadvantage of PQ scheduling is that when congestion occurs, if messages exist in high priority queues for a long time, messages in low priority queues will not get scheduling opportunities.

- WRR Scheduling

WRR(Weighted Round Robin) scheduling is the weighted polling scheduling. WRR performs alternate scheduling among queues to ensure that each queue gets a certain service time.

Taking the interface with eight output queues as an example, WRR configures a weighted value (w7, w6, w5, w4, w3, w2, w1, w0 in turn) for each queue, and the weighted value indicates the proportion of acquired resources. For a more specific example, for a 100M interface, the weighted values of its WRR algorithm are configured as 50, 50, 30, 30, 10, 10, 10, 10 (corresponding to w7, w6, w5, w4, w3, w2, w1, w0 in turn), which can ensure that the lowest priority queue obtains at least 5M bandwidth and avoid

WRR has another advantage: although multiple queues are scheduled in turn, service time slices are not fixed for each queue, that is to say, if one queue is empty, immediately switch to the next queue for scheduling, so that bandwidth resources can be fully utilized.

WRR scheduling has two disadvantages:

- WRR scheduling is based on the number of messages, while users are generally concerned about bandwidth. When the average message length of each queue is equal or known, users can obtain the desired bandwidth by configuring WRR weight; However, when the average message length of the queue changes, users cannot obtain the desired bandwidth by configuring WRR weights.
- Delay-sensitive services (such as voice) cannot be scheduled in time.

- WDRR Scheduling

The implementation principle of WDRR(Weighted Deficit Round Robin) scheduling is basically the same as that of WRR scheduling.

The difference between WDRR scheduling and WRR scheduling is that WRR scheduling is based on the number of messages, while WDRR scheduling is based on the length of messages. If the message length exceeds the scheduling capacity of the queue, WDRR scheduling allows negative weights to ensure that long messages can also be scheduled. However, the queue will not be scheduled at the next polling schedule, and will not participate in WDRR scheduling until the weight is positive.

WDRR scheduling avoids the disadvantage that messages in low priority queues can not be served for a long time when congestion occurs and PQ scheduling is adopted, and also avoids the disadvantage that WRR scheduling cannot allocate bandwidth resources according to the allocation ratio when the message lengths of each queue are unequal or change greatly.

However, WDRR scheduling also has the disadvantage that delay-sensitive services (such as voice) cannot be scheduled in time.

When all queues participating in WDRR scheduling have the same weight, WDRR scheduling has the same effect as DRR scheduling.

- WFQ Scheduling

The purpose of Fair Queue (FQ) is to share network resources as fairly as possible, to optimize the delay and jitter of all flows, and to give different queues fair scheduling opportunities. WFQ(Weighted Fair Queue) scheduling increases the priority consideration on the basis of FQ, so that high-priority messages have more chances to get priority scheduling than low-priority messages.

WFQ can automatically classify streams according to the "session" information of streams (protocol type, source and destination TCP or UDP port numbers, source and

destination IP addresses, priority bits in ToS domain, etc.), and provide as many queues as possible, so as to evenly put each stream into different queues, thus balancing the delay of each stream as a whole. When leaving the queue, WFQ allocates the bandwidth that each flow should occupy at the exit according to the priority of the flow. The smaller the value of priority, the less bandwidth is obtained. The larger the value of priority, the more bandwidth is obtained.

- PQ+WRR/PQ+WDRR/PQ+WFQ Scheduling

PQ scheduling and WRR/WDRR/WFQ scheduling have their own advantages and disadvantages. When PQ scheduling is used alone, messages in low priority queues will not get bandwidth for a long time, while low-latency services will not get priority scheduling when WRR/WDRR/WFQ scheduling is used alone. PQ+WRR/PQ+WDRR/PQ+WFQ scheduling method combines the former two scheduling methods, which can not only give full play to the advantages of the two scheduling methods, but also overcome their respective shortcomings.

Users can use PQ+WRR/PQ+WDRR/PQ+WFQ scheduling mode to put important protocol messages and delay-sensitive service messages into queues scheduled by PQ, and allocate specified bandwidth for the queues. Other messages are put into each queue scheduled by WRR/WDRR/WFQ according to their respective priorities, and each queue is cyclically scheduled according to the weights.

25.2 QoS Configuration

25.2.1 Configure Global QOS Enable/Disable

【Command】

```
mls qos enable
no mls qos
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

On is enable, disable is no

【Description】

For configuring global QOS on and off, the MLS QOS switch must be on for all qos configurations.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mls qos enable
Switch(config)#no mls qos
```

25.2.2 Configure the Queue Bitmap

【Command】

```
mls qos cos-map <0-7> <0-7> <0-7> <0-7> <0-7> <0-7> <0-7>  <0-7>
no mls qos cos-map
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

Parameter 1: select a queue for COS 0 message;

Parameter 2: select a queue for COS 1 message;

Parameter 3: select a queue for COS 2 message;

Parameter 4: select a queue for COS 3 message;

Parameter 5: select a queue for COS 4 message;

Parameter 6: select a queue for COS 5 message;

Parameter 7: select a queue for COS 6 message;

Parameter 8: select a queue for COS 7 message;

【Description】

Configure the queue value for each COS. No Delete.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mls qos enable
Switch(config)#mls qos cos-map 1 2 3 4 5 6 7 0
Switch(config)#no mls qos cos-map
```

25.2.3 Configure Queue Scheduling Mode

【Command】

mls qos scheduler (sp|wrr <1-10> <1-10> <1-10> <1-10> <1-10> <1-10> <1-10> <1-10> <1-10>)

no mls qos scheduler

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

sp: represents strict priority; wrr means Weighted Round Robin, configuring each queue with a weight according to weight priority <1-10>.

no: means delete. The default mode is simple polling mode SSR.

【Description】

SP: Strict Priority, the SP schedule sends packets in the higher-priority queue in Strict Priority order from highest to lowest, and then sends packets in the lower-priority queue when the higher-priority queue is empty. Queue 7 has the highest priority and queue 0 has the lowest priority.

WRR, weighted scheduling based on message, can configure how many messages are scheduled per queue as possible and then transfer to the next queue.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mls qos enable
Switch(config)#mls qos schedule sp // configured to SP mode
Switch(config)#mls qos schedule wrr 1 2 3 4 5 6 7 1 // the configuration effect is: each queue takes away the message with the corresponding weight ratio
Switch(config)#no mls qos schedule // restore default configuration SRR
```

25.2.4 Configure the DSCP-COS Bitmap

【Command】

```
mls qos map dscp-cos NAME (<0-63>|<0-63> <0-63>|<0-63> <0-63>
<0-63>|<0-63> <0-63> <0-63> <0-63>|<0-63> <0-63> <0-63> <0-63>
<0-63>|<0-63> <0-63> <0-63> <0-63> <0-63> <0-63>|<0-63> <0-63>
<0-63> <0-63> <0-63> <0-63> <0-63>|<0-63> <0-63> <0-63> <0-63>
<0-63> <0-63> <0-63> <0-63>) to <0-7>
no mls qos map dscp-cos (NAME)
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

NAME: create a name for the dscp-cos map.

<0-63> to <0-7>: transfer each DSCP value to the corresponding COS queue.

No means to delete.

【Description】

The default dscp-cos map is 0-7 to cos 0 8-15 to cos 1 16-23 to cos 2 24-31 to cos 3 32-39 to cos 4 40-47 to cos 5 48-55 to cos 6 56-63 to cos 7.

The configuration is only issued when it is referenced. No means to delete NAME, specifying which one to delete.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mls qos enable
Switch(config)# mls qos map dscp-cos dscp1 10 46 56 to 6// means
that the message DSCP value is 10 46 56 and so on will be transfered
to queue 6
Switch(config)#no mls qos map dscp-cos dscp1// specify to delete
dscp1
```

25.2.5 Configure DSCP -DSCP Bitmap

【Command】

```
mls qos map dscp-mutation NAME (<0-63>|<0-63> <0-63>|<0-63> <0-63>
<0-63>|<0-63> <0-63> <0-63> <0-63>|<0-63> <0-63> <0-63> <0-63>
<0-63>|<0-63> <0-63> <0-63> <0-63> <0-63> <0-63>|<0-63> <0-63>
<0-63> <0-63> <0-63> <0-63> <0-63>|<0-63> <0-63> <0-63> <0-63>
<0-63> <0-63> <0-63> <0-63>) to <0-63>
no mls qos map dscp-mutation （NAME）
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

By default, it is dscp-mutation map 0-63 to dscp 0-63

No means to delete, name is to specify which MAP to delete, all means to delete all.

【Description】

When different DSCP values received, some changes can be made to the DSCP and then transfer to different queues.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mls qos map dscp-mutation dscp2  22 61 to 2
Switch(config)#no mls qos map dscp-mutation dscp2
```

25.2.6 Create a CLASS-MAP

【Command】

```
class-map NAME
no class-map NAME
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

No means to delete.

【Description】

Class map: Class map is a definition of a Class map that groups different types of data flows.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#class-map ac
Switch(config-class-map-ac)#no class-map ac
```

25.2.7 Create a POLICY-MAP

【Command】

```
policy-map NAME
no policy-map NAME
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

No means to delete.

【Description】

policy map: It is a definition of a policy map that matches a class map to determine the bandwidth and/or priority of a class of data flows.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#policy-map ad
Switch(config-pmap_ad)#no policy-map ad
```

25.2.8 Configure the CLASS-MAP Property

【Command】

```
match access-group NAME
match ip-dscp (<0-63>|<0-63> <0-63>|<0-63> <0-63> <0-63>|<0-63>
<0-63> <0-63> <0-63>|<0-63> <0-63> <0-63> <0-63> <0-63>|<0-63>
<0-63> <0-63> <0-63> <0-63> <0-63>|<0-63> <0-63> <0-63> <0-63>
<0-63> <0-63> <0-63>|<0-63> <0-63> <0-63> <0-63> <0-63> <0-63>
<0-63> <0-63>)
no match ip-dscp

match ip-precedence (<0-7>|<0-7> <0-7>|<0-7> <0-7> <0-7>|<0-7>
<0-7> <0-7> <0-7>|<0-7> <0-7> <0-7> <0-7> <0-7>|<0-7> <0-7> <0-7>
<0-7> <0-7> <0-7>|<0-7> <0-7> <0-7> <0-7> <0-7> <0-7> <0-7>|<0-7>
<0-7> <0-7> <0-7> <0-7> <0-7> <0-7> <0-7>)
no match ip-precedence <0-7>

match layer4 (source-port|destination-port) <1-65535>
no match layer4 (source-port|destination-port) <1-65535>

match vlan <1-4094>
match vlan-range <1-4094> to <1-4094>
no match vlan
```

【View】

CLASS-MAP Configuration View

【Default Level】

2: Configuration level

【Parameter】

ip-dscp: matches dscp value of IP message

ip-precedence: match the priority value of dscp of IP message

layer4: Match L4 port number

vlan: Match vlan id value

【Description】

None

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#class-map ac
```

```
*Switch(config-class-map_ac)#match layer4 destination-port 80
*Switch(config-class-map_ac)#no match layer4 destination-port 80
```

25.2.9 Configure the POLICY-MAP Property

【Command】

```
class NAME
```

【View】

Policy configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

Associate classes in the policy configuration mode and enter the policy class configuration mode

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mls qos enable
*Switch(config)#policy-map ac
*Switch(config-pmap_ac)#class aa
*Switch(config-pmap-ac-class_aa)#
```

25.2.10 Configure the POLICY-MAP-C Property

【Command】

```
set cos （<0-7>|cos-inner）
no set cos

set ip-dscp <0-63>
no set ip-dscp

set ip-precedence <0-7>
no  set ip-precedence

police <64-1000000> <4-20000000> exceed-action drop
```

no police <64-1000000> <4-20000000> exceed-action drop

【View】

Policy class configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

COS setting priority, and no means to restore the default.

IP-DSCP: set dscp value of IP message

IP-priority: set the value of IP message priority

Police < 64-1000000 > < 4-2000000 > exceed-action drop: limit the bandwidth of the matched bound flow and discard the messages with the bandwidth exceeding the limit, and the rate should be a multiple of 64.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#mls qos enable
*Switch(config)#policy-map ad
*Switch(config-pmap_ad)#class ac
*Switch(config-pmap_ad-class_ac)#police 2048 2048 exceed-action
drop
```

25.2.11 Configure QOS Interface Mode

【Command】

```
service-policy input NAME
no service-policy input NAME

mls qos trust dscp
no mls qos trust dscp

mls qos cos <0-7>
no mls qos cos

mls qos dscp-cos NAME
no mls qos dscp-cos NAME
```

mls qos dscp-mutation NAME
no mls qos dscp-mutation NAME

【View】

Ethernet port configuration view

【Default Level】

2: Configuration level

【Parameter】

Name

【Description】

service-policy input NAME: means to install the contents of the policy-map into the specified interface.

No means cancel the installation.

mls qos trust dscp: means that messages received at the specified port are queued according to the value of DSCP, and the default mode is COS. No means restore to default value.

mls qos dscp-cose NAME: installs the named dscp-mutation map to the specified port. No means to delete.

mls qos dscp-mutation NAME: installs the named dscp-mutation map to the specified port. No means to delete.

mls qos cos <0-7>: configure the default priority of the specified port. No means to restore the default value, which is 0.

【Instance】

```
Switch> enable
Switch#configure terminal
*Switch(config)#interface ge1
*Switch(config-ge1)#service-policy input ad
*Switch(config-ge1)#exit
*Switch(config)#show policy-map
  POLICY-MAP-NAME: ad
    State: attached
    CLASS-MAP-NAME: ac
      Police: average rate (2048 kbps)
              burst size (2048 bytes)
              exceed-action (drop)
              excess burst size (2048 bytes)
              flow control mode (none)
```

```
*Switch(config)#show mls qos interface ge1
  INPUT-POLICY-MAP-NAME: ad
    CLASS-MAP-NAME: ac
      Police: average rate (2048 kbps)
             burst size (2048 bytes)
             exceed-action (drop)
             excess burst size (2048 bytes)
             flow control mode (none)
  Trust Mode: Ports default priority
  Port Default Prioiry: 0
  VLAN Priority Overide: Not Configured
  Egress Traffic Shaping: Not Configured
View all information on configuring MLS qos within a port.
```

25.2.12 Display the Queue Bitmap

【Command】

show mls qos cos-map

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

Check the queue value for each COS.

【Instance】

```
Switch> enable
Switch# show mls qos cos-map
Cos-queue map:
       COS  :   0   1   2   3   4   5   6   7
       ---------------------------------------
       QUEUE:   1   2   3   4   5   6   7   0
```

25.2.13 Display Queue Scheduling Mode

【Command】

show mls qos scheduler

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

sp: represents strict priority; wrr means Weighted Round Robin, configuring each queue with a weight according to weight priority <1-10>.

Show: to view the configuration.

【Description】

SP: Strict Priority, the SP schedule sends packets in the higher-priority queue in Strict Priority order from highest to lowest, and then sends packets in the lower-priority queue when the higher-priority queue is empty. Queue 7 has the highest priority and queue 0 has the lowest priority.

WRR, weighted scheduling based on message, can configure how many messages are scheduled per queue as possible and then transfer to the next queue.

【Instance】

```
Switch> enable
Switch# show mls qos schedule
 Strict Priority
```

25.2.14 Display DSCP-COS Bitmap

【Command】

show mls qos maps **dscp-cos (NAME)**

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

NAME: create a name for the dscp-cos map.

Show: means to view the configuration.

【Description】

The default dscp-cos map is 0-7 to cos 0 8-15 to cos 1 16-23 to cos 2 24-31 to cos 3 32-39 to cos 4 40-47 to cos 5 48-55 to cos 6 56-63 to cos 7.

【Instance】

```
Switch> enable
Switch# show  mls qos maps dscp-cos dscp1
  DSCP-TO-COS-MAP:  dscp1
    d1 :   d2  0   1   2   3   4   5   6   7   8   9
    ----------------------------------------------
     0 :       0   0   0   0   0   0   0   0   1   1
     1 :       6   1   1   1   1   1   2   2   2   2
     2 :       2   2   2   2   3   3   3   3   3   3
     3 :       3   3   4   4   4   4   4   4   4   4
     4 :       5   5   5   5   5   5   6   5   6   6
     5 :       6   6   6   6   6   6   6   7   7   7
     6 :       7   7   7   7
```

25.2.15 Display DSCP- DSCP Bitmap

【Command】

```
show mls qos maps dscp-mutation(NAME)
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

NAME means to specify which MAP to view, all means to view all.

【Description】

Show: means to view.

【Instance】

```
Switch> enable
Switch# show mls qos maps dscp-mutation dscp2
```

```
DSCP-TO-DSCP-MAP:  dscp2
   d1 :   d2  0   1   2   3   4   5   6   7   8   9
   ----------------------------------------------
    0 :       0   1   2   3   4   5   6   7   8   9
    1 :      10  11  12  13  14  15  16  17  18  19
    2 :      20  21   2  23  24  25  26  27  28  29
    3 :      30  31  32  33  34  35  36  37  38  39
    4 :      40  41  42  43  44  45  46  47  48  49
    5 :      50  51  52  53  54  55  56  57  58  59
    6 :      60   2  62  63
```

25.2.16 Display a CLASS-MAP

【Command】

show class-map(NAME|)

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

Show: means to view the specified class-map or all class-maps including the information configured therein.

【Description】

Class map: Class map is a definition of a Class map that groups different types of data flows.

【Instance】

```
Switch> enable
Switch# show class-map
CLASS-MAP-NAME:ac
```

25.2.17 Display a POLICY-MAP

【Command】

show policy-map (NAME|)

【View】

Priviledged user mode

【Default Level】

2: view level

【Parameter】

show: means to view the information specified or all included in it.

【Description】

policy map: It is a definition of a policy map that matches a class map to determine the bandwidth and/or priority of a class of data flows.

【Instance】

```
Switch> enable
Switch# show policy-map ad
POLICY-MAP-NAME:ad
  State:attached
```

26 ACL Configuration

26.1 Overview

The ACL(Access Control List) is a set composed of one or more rules. Rule refers to the judgment statement describing the message matching condition. These conditions may be the source address, destination address, port number of message.

ACL is essentially a message filter, and rule is the filter element of the filter. The device matches messages based on these rules, which can filter out specific messages and allow or prevent the messages from passing according to the processing strategy of the service module applying ACL.

With the rapid development of network, the problems of network security and quality of service (QoS) have become increasingly prominent.

- Important server resources of enterprises are randomly accessed, and confidential information of enterprises is easy to be leaked, resulting in potential safety hazards.
- Internet viruses invade the intranet wantonly, and the security of intranet environment is worrying.
- The network bandwidth is arbitrarily occupied by various services, and the bandwidth of voice and video services with the highest service quality requirements cannot be guaranteed, resulting in poor user experience.

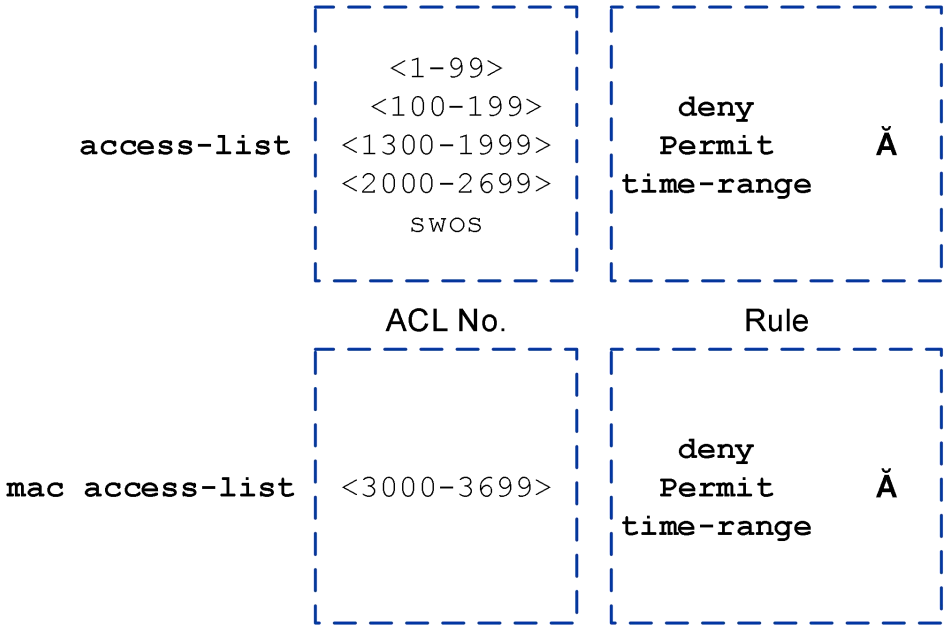
All these problems have a great impact on the normal network communication. Therefore, it is extremely urgent to improve the network security service quality. In this case, ACL came into being.

ACL can realize accurate identification and control of message flow in the network, and achieve the purpose of controlling network access behavior, preventing network attacks and improving network bandwidth utilization, thus ensuring the security of network environment and the reliability of network service quality.

26.2 Principles

26.2.1 ACL Principles

ACL consists of a series of rules. By matching messages with ACL rules, devices can filter out specific messages.



- ACL number: used to identify ACL.
  According to the different functions of ACL rules, ACLs are divided into standard ACLs, extended ACLs and layer 2 ACLs, and the value range of each type of ACL number is different.
  - Standard ACL: Rules are made only based on the source IP address.
  - Expand ACL: make rules according to layer 3 or layer 4 information such as source IP address information, destination IP address information, protocol type and protocol characteristics of IP carrier.
  - Layer 2 ACL: Rules are made according to layer 2 information such as source MAC address and destination MAC address.

  In addition to identifying ACLs by ACL numbers, devices also support identifying ACLs by SWOS names, just like replacing IP addresses with domain names, which is more convenient to remember. This kind of ACL is called named ACL.
- Rule: judgment statements describing the matching conditions of messages.
  - Action: including permit/deny, which means allow/deny.
  - Matches: ACL defines extremely rich matches. In addition to source address and effective time period, ACL also supports many other rule

matches. For example, Layer 2 Ethernet header information (such as source MAC, destination MAC and Ethernet frame protocol type), Layer 3 message information (such as destination address and protocol type), and Layer 4 message information (such as TCP/UDP port number), etc.

26.2.1.1 Matching Mechanism of ACL

When a device matches a message with ACL rules, it follows the mechanism of "stop matching once it hits".

First, the system will find out if ACL is configured on the device.

- If the ACL does not exist, the ACL matching result is: no match.
- If an ACL exists, the system will start with the rule with the lowest number in the ACL.
  - If the permit rule is matched, stop looking for the rule and return the ACL matching result as: match (allow).
  - If the deny rule is matched, stop looking for the rule and return the ACL matching result as: match (reject).
  - If the previous rule is not matched, continue to find the next rule, and loop. If the last rule is found all the time and the message is still not matched, the ACL matching result is no match.

From the whole ACL matching process, it can be seen that there are two matching results: "match" and "mismatch".

- Match (hit rule): refers to the existence of ACL, and the rule matching the matching condition is found in ACL.
  No matter whether the matching action is "permit" or "deny", it is called "matching", not just matching the permit rule.
- Unmatched (missed rule): refers to the fact that there is no ACL, or there are no rules in the ACL, or all the rules in the ACL are traversed, but no matching rule is found.

26.2.2 Classification of ACL

26.2.2.1 Partition of Identification Method Based on ACL

Divided as follows:

- Digital ACL: Traditional ACL identification method. When creating an ACL, specify a unique number to identify the ACL.
- Name ACL: ACL is identified by name instead of number.

Users can assign numbers to ACLs when creating them, and different numbers correspond to different types of ACLs. At the same time, in order to facilitate memory and identification, users can also create named ACLs, that is, when creating ACLs, set their names.

26.2.2.2 Division Based on ACL Rule Definition

Rule definition methods based on ACL are divided as follows.

Classify	Rule definition description	Number range
Standard ACL	Only the source IP address, fragmentation information and effective time period information of the message are used to define the rule.	1-99、1300-1999
Extended ACL	Rules can be defined by using not only the source IP address of IPv4 message, but also the destination IP address, IP protocol type, ICMP type, TCP source/destination port, UDP source/destination port number, effective time period, etc.	100-199、2000-2699
Layer 2 ACL	Rules are defined by Ethernet header information of messages, such as source MAC(Media Access Control) address, destination MAC address, etc.	3000-3699

26.2.3 Common Matches of ACL

There are many kinds of ACL matches supported by devices, among which the most commonly used matches include the following.

26.2.3.1 Effective Time Period

Format: **`time-range TIME-NAME`**

All ACLs support filtering messages according to the effective time period. See "ACL Effective Time Period" in the following subsection for details.

26.2.3.2 Protocol Type Carried by IP

Format: **<0-255> | eigrp | esp | gre | icmp | igmp | ip | ipinip | ospf | pcp | pim | rsvp | tcp | udp | vrrp**

Extended ACL supports message filtering based on protocol type. Commonly used protocol types include: ICMP (protocol No.1), TCP (protocol No.6), UDP (protocol No.17), GRE (protocol No.47), IGMP (protocol No.2), IP (any IP layer protocol), IPinIP (protocol No.4), OSPF (protocol No.89). 0-255 indicates the value of the protocol number.

26.2.3.3 Source/Destination IP Address and Its Wildcard Mask

The format of the source IP address and its wildcard mask: **（A.B.C.D A.B.C.D） | any | (host A.B.C.D)**

The format of the destination IP address and its wildcard mask: **（A.B.C.D A.B.C.D） | any | (host A.B.C.D)**

Standard ACL supports filtering messages according to source IP address, while extended ACL supports filtering messages according to not only source IP address but also destination IP address.

When defining a source/destination IP address as a rule match, it is necessary to specify a wildcard mask after the source/destination IP address field to determine an address range together with the source/destination IP address field.

The wildcard mask of IP address is similar to the reverse subnet mask of IP address, and it is also a 32-bit numeric string used to indicate which bits in IP address will be checked. In each bit, "0" means "check the corresponding bit", "1" means "don't check the corresponding bit", which can be summarized as "check 0, ignore 1". However, unlike IP address subnet mask, "0" and "1" in subnet mask must be continuous, while "0" and "1" in wildcard mask can be discontinuous.

Wildcard mask can be 0, which is equivalent to 0.0.0.0, indicating that the source/destination address is the host address; It can also be 255.255.255.255, which means any IP address, which is equivalent to specifying **any** parameter.

26.2.3.4 Source/Destination MAC Address and Its Wildcard Mask

The format of the source MAC address and its wildcard mask: **（MAC MASK） | any | (host MAC)**

The format of the destination address and its wildcard mask: **（MAC MASK）| any | (host MAC)**

Only layer 2 ACL supports filtering messages based on source/destination MAC address.

When defining a source/destination MAC address as a rule match, a wildcard mask can be specified at the same time after the source/destination MAC address field, which can be used to determine an address range together with the source/destination MAC address field.

The format of the MAC address wildcard mask is the same as that of the MAC address, which is represented by hexadecimal number and consists of six bytes (48 bits), which is used to indicate which bits in the MAC address will be checked. Different from IP address wildcard mask, in each bit of wildcard mask of MAC address, 1 means "check corresponding bit" and 0 means "don't check corresponding bit". The mask is ffff-ffff-ffff, which means that every bit of the MAC address is the host MAC; The mask is 0000-0000-0000, which means any MAC address. If no wildcard mask is specified, the default mask is ffff-ffff-ffff, which means to check every bit of the MAC address.

26.2.3.5 TCP/UDP Port Number

The format of source port number: (**eq|gt|lt|ne <0-65535>)|(rang <0-65535> <0-65535>)**

The format of destination port number: (**eq|gt|lt|ne <0-65535>)|(rang <0-65535> <0-65535>)**

In the extended ACL, when the protocol type is specified as TCP or UDP, the device supports filtering messages based on the source/destination port number of TCP/UDP.

The meaning of the comparison operators of TCP/UDP port number is as follows:

- **eq <0-65535>:** specify equal to source/destination port.
- **gt <0-65535>:** specify larger than the source/destination port.
- **lt <0-65535>**: specify less than the source/destination port.
- **ne <0-65535>**:specify not equal to the source/destination port.
- **rang <0-65535> <0-65535>**: specify the range of source/destination ports. **<0-65535> <0-65535>** respectively indicate the beginning and end of the port range.

TCP/UDP port numbers can be expressed by numbers or strings (mnemonics). For example, **access-list 100 deny tcp any any eq 80** can be replaced by **access-list 100 deny tcp any any eq www**.

Common TCP port numbers and corresponding strings are shown in the following table.

Port	Character string	Protocol	Note
20	ftp-data	FTP data connections	FTP data port
21	ftp	File Transfer Protocol(FTP)	File transfer protocol (FTP) port
23	telnet	Telnet	Telnet service
25	smtp	Simple Mail Transport Protocol (SMTP)	Simple Mail Transport Protocol
80	www	World Wide Web (HTTP)	Hyper Text Transfer Protocol (HTTP) for World Wide Web (WWW) services for web browsing
110	pop3	Post Office Protocol v3	Mail protocol-version 3

Common UDP port numbers and corresponding strings are shown in the following table.

Port	Character string	Protocol	Note
69	tftp	Trivial File Transfer Protocol (TFTP)	Trivial File Transfer Protocol
161	snmp	SNMP	Simple Network Management Protocol
162	snmptrap	SNMPTRAP	SNMP Trap
520	rip	Routing Information Protocol	RIP routing protocol

26.2.4 Effective Time Period of ACL

26.2.4.1 Background

ACL defines abundant matches, which can meet most of the message filtering requirements. But the demand is constantly changing and developing, and new

demands are always emerging. For example, a company requires that employees are only allowed to browse several websites related to work during working hours, and other Internet websites can be accessed after work or weekends; For another example, in the peak period of network traffic from 20: 00 to 22: 00 every day, in order to prevent P2P and download services from occupying a large amount of bandwidth and affecting the normal use of other data services, it is necessary to limit the bandwidth of P2P and download services.

ACL filtering based on time is used to solve the above problems. Administrators can configure one or more ACL effective time periods according to the requirements of network access behavior and network congestion, and then refer to the time periods in ACL rules, so as to set different policies in different time periods and achieve the purpose of network optimization.

26.2.4.2 Effective Time Period Mode

There are two modes for the effective time period referenced in ACL rules:

- The first mode--cycle time period: the time range is defined by taking the week as the parameter, which means that the rule takes effect cyclically with a week cycle (e.g., 8: 00 to 12: 00 every Monday).

  Format: **`perioodic HH:MM:SS to HH:MM:SS (DAY | daily | offday | workday)`**

  - **`HH:MM:SS to HH:MM:SS`**: start time and end time. The format is "hours: minutes: seconds" to "hours: minutes: seconds".
  - **`DAY`**: value range < 0-6 >, 0 means Sunday, 1 means Monday, … 6 means Saturday. DAY combination of one or severals.
  - **`workday`**: five days from Monday to Friday.
  - **`daily`**: It includes seven days a week.
  - **`offday`**: including Saturday and Sunday, two days.
- The second mode--absolute time period: it starts from a certain time on a certain day of a certain year and ends at a certain time on a certain day of a certain year, which means that the rules will take effect within this time range.

  Format: **`absolute start HH:MM:SS YYYY-MM-DD end HH:MM:SS YYYY-MM-DD`**

  - **`start HH:MM:SS YYYY-MM-DD`**: start time "hour: minute: second year-month-day".
  - **`end HH:MM:SS YYYY-MM-DD`**: end time "hour: minute: second year-month-day".

User can use the same name (**time-range NAME**) to configure multiple time periods with different contents, and the intersection between configured periodic time periods and absolute time periods will become the final effective time range.

26.3 ACL Configuration

26.3.1 Configure IPv4 Extended ACL Based on IP Addresses

【Command】

access-list (<1-99>|<1300-1999>) (deny|permit) ((A.B.C.D [A.B.C.D]) **| (host** A.B.C.D**) | any)**

no access-list (<1-99>|<1300-1999>) (deny|permit) ((A.B.C.D [A.B.C.D]) **| (host** A.B.C.D**) | any)**

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

(<1-99>|<1300-1999>) : represents the scope of the standard ACL.

(deny|permit) : ACL action, deny, permit.

(A.B.C.D [A.B.C.D]): represents the source IP address and mask. The mask adopts the anti-code mechanism, such as 192.168.1.1 0.0.0.0 means only match 192.168.1.1 source IP message.

host A.B.C.D: indicates that the source IP address is A.B.C.D and the mask is 0.0.0.0.

any: indicates that the source IP address and mask are 0.0.0.0 255.255.255.255, which means all IP addresses.

【Description】

Access-list: command is used to create a standard filter rule group. A group can support up to 32 rules. No is to delete a rule group. When the message matches the corresponding rule, the action will be executed. For example, the configuration rule is as follows: access-list 1 deny 192.168.1.1 0.0.0.0.When the message from 192.168.1.1 is received, the action performed is discarded. These rules only take effect when activated on a port using the ip-access-group command. And has the "first activation first effect" feature.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 1 deny 192.168.1.1 0.0.0.0
```

26.3.2 Configure IPv4 Extended ACL Based on IP Addresses

【Command】

```
access-list (<100-199>|<2000-2699>) (deny|permit) ip ((A.B.C.D A.B.C.D) | (host A.B.C.D) | any) ((A.B.C.D A.B.C.D) | (host A.B.C.D) | any)
no access-list (<100-199>|<2000-2699>) (deny|permit) ip ((A.B.C.D A.B.C.D) | (host A.B.C.D) | any) ((A.B.C.D A.B.C.D) | (host A.B.C.D) | any)
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

(<100-199>|<2000-2699>) : indicates the scope of the extended ACL.

(deny|permit) : ACL action, deny, permit.

((A.B.C.D A.B.C.D) | (host A.B.C.D) | any) ((A.B.C.D A.B.C.D) | (host A.B.C.D) | any): The former represents the source IP address and mask information, while the latter represents the destination IP address and mask information. The specific information is as follows.

- (A.B.C.D A.B.C.D): represents the source/destination IP address and mask. The mask adopts the anti-code mechanism. For example, 192.168.1.1 0.0.0.0, which means that only packets match 192.168.1.1 source/destination IP.
- host A.B.C.D: indicates that the source/destination IP address is A.B.C.D and the mask is 0.0.0.0.
- any: indicates that the source /destination IP address is 0.0.0.0 255.255.255.255, which means all IP addresses.

No means to delete the corresponding rule.

【Description】

Access-list: command is used to create a standard filter rule group. A group can support up to 32 rules. No is to delete a rule group. When the message matches the corresponding rule, the action will be executed. For example, the configuration rule is

as follows: access-list 101 deny 192.168.1.1 0.0.0.0 192.168.2.1 0.0.0.0 When the message from 192.168.1.1 to 192.168.2.1 is received, the action performed is discarded. These rules only take effect when activated on a port using the ip-access-group command. And has the "first activation first effect" feature.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 101 deny ip 192.168.1.1 0.0.0.0
192.168.2.1 0.0.0.0
```

26.3.3 Configure Other IPv4 Protocol Extended ACL based on IP Addresses

【Command】

```
access-list (<100-199>|<2000-2699>) (deny|permit)
(<0-255>|eigrp|esp|gre|icmp|igmp|ipinip|ospf|pcp|pim|rsvp|vrrp)
((A.B.C.D A.B.C.D)|(any)|(host A.B.C.D)) ((A.B.C.D
A.B.C.D)|(any)|(host A.B.C.D))
no access-list (<100-199>|<2000-2699>) (deny|permit)
(<0-255>|eigrp|esp|gre|icmp|igmp|ipinip|ospf|pcp|pim|rsvp|vrrp)
((A.B.C.D A.B.C.D)|(any)|(host A.B.C.D)) ((A.B.C.D
A.B.C.D)|(any)|(host A.B.C.D))
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

(<100-199>|<2000-2699>) : indicates the scope of the extended ACL.

(deny|permit) : ACL action, deny, permit.

(<0-255>|eigrp|esp|gre|icmp|igmp|ipinip|ospf|pcp|pim|rsvp|vrrp): configure ip protocol type:

- <0-255>：An IP protocol number
- eigrp：EIGRP routing protocol
- esp：Encapsulation Security Payload
- gre：General Routing Encapsulation
- icmp：Internet Control Message Protocol
- igmp：Internet Group Managent Protocol

- ipinip：iP in IP tunneling
- ospf：OSPF routing protocol
- pcp：Payload Compression Protocol
- pim：Protocol Independent Multicast
- rsvp：Resource Reservation Protocol
- vrrp：Virtual Router Redundancy Protocol

((A.B.C.D A.B.C.D)|(any)|(host A.B.C.D)) ((A.B.C.D A.B.C.D)|(any)|(host A.B.C.D)): The former represents the source IP address and mask information, while the latter represents the destination IP address and mask information. The specific information is as follows.

- (A.B.C.D A.B.C.D): represents the source/destination IP address and mask. The mask adopts the anti-code mechanism. For example, 192.168.1.1 0.0.0.0, which means that only packets match 192.168.1.1 source/destination IP.
- host A.B.C.D: indicates that the source/destination IP address is A.B.C.D and the mask is 0.0.0.0.
- any: indicates that the source /destination IP address is 0.0.0.0 255.255.255.255, which means all IP addresses.

No means to delete.

【Description】

Since the message has a corresponding protocol port number, it can be configured to filter based on the protocol port number. For example, the configuration rules are as follows: access-list 101 deny ahp 192.168.1.1 0.0.0.0 192.168.2.1 0.0.0.0.When the pim message from 192.168.1.1 to 192.168.2.1 is received, the action performed is discarded. These rules only take effect when activated on a port using the ip-access-group command. And has the ″first activation first effect” feature.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 101 deny pim 192.168.1.1 0.0.0.0
192.168.2.1 0.0.0.0
```

26.3.4 Configure IPv4 TCP Extended ACL Based on IP Addresses

【Command】

```
access-list (<100-199>|<2000-2699>) (deny|permit) (tcp) ((A.B.C.D
A.B.C.D)|(any)|(host A.B.C.D)) [eq
(<1-65535>|ftp|ftp-data|pop3|smtp|telnet|www)] ((A.B.C.D A.B.C.D)
```

|(any)|(host A.B.C.D**))** **[eq**
(<1-65535>**|ftp|ftp-data|pop3|smtp|telnet|www)]**
[ack|fin|psh|rst|syn|urg]
no access-list **(**<100-199>|<2000-2699>**)** **(deny|permit)** **(tcp)**
((A.B.C.D A.B.C.D**)|(any)|(host** A.B.C.D**))** **[eq**
(<1-65535>**|ftp|ftp-data|pop3|smtp|telnet|www)]** **((**A.B.C.D A.B.C.D**)**
|(any)|(host A.B.C.D**))** **[eq**
(<1-65535>**|ftp|ftp-data|pop3|smtp|telnet|www)]**
[ack|fin|psh|rst|syn|urg]

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

(<100-199>|<2000-2699>) : indicates the scope of the extended ACL.

(deny|permit) : ACL action, deny, permit.

Tcp: filter tcp protocol message.

((A.B.C.D A.B.C.D)|(any)|(host A.B.C.D))…((A.B.C.D A.B.C.D)|(any)|(host A.B.C.D)): The former represents the source IP address and mask information, while the latter represents the destination IP address and mask information. The specific information is as follows.

- (A.B.C.D A.B.C.D): represents the source/destination IP address and mask. The mask adopts the anti-code mechanism. For example, 192.168.1.1 0.0.0.0, which means that only packets match 192.168.1.1 source/destination IP.
- host A.B.C.D: indicates that the source/destination IP address is A.B.C.D and the mask is 0.0.0.0.
- any: indicates that the source /destination IP address is 0.0.0.0 255.255.255.255, which means all IP addresses.

eq: specify equal to source/destination port.

(<1-65535>|ftp|ftp-data|pop3|smtp|telnet|www): corresponding to different TCP message types:

- Port number(1-65535),
- File Transfer Protocol (21),
- FTP data connections (20),
- Post Office Protocol v3 (110),
- Simple Mail Transport Protocol (25),
- Telnet (23),
- World Wide Web (HTTP, 80)。

[ack|fin|psh|rst|syn|urg]: The device supports filtering messages based on TCP flag information. The TCP header has 6 flag bits, and the ACL rule specifying TCP flag information can be used to realize one-way access control.

- Match on the Ack bit, which indicates that the serial number is valid;
- Match on the FIN bit, which indicates that the sender completes the sending task;
- Match on the Psh bit, which indicates that the receiver should submit this message segment to the application layer as soon as possible;
- Match on the Rst bit, which identifies the reestablished connection;
- Match on the Syn bit, which is used to initiate a connection;
- Match on the Urg bit, which indicates that the urgent pointer is valid.

No means to delete.

【Description】

Configure extended ACL TCP protocol based on IPv4. For example, the configuration rule is as follows: access-list 101 deny tcp host 192.168.1.1 eq ftp host 192.168.2.1 eq pop3 fin, then when receiving the tcp message from 192.168.1.1 to 192.168.2.1, the action executed is discard. These rules only take effect when activated on a port using the ip-access-group command. And has the "first activation first effect" feature.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 101 deny tcp host 192.168.1.1 eq ftp
host 192.168.2.1 eq pop3 fin
```

26.3.5 Configure IPv4 UDP Extended ACL Based on IP Addresses

【Command】

```
access-list (<100-199>|<2000-2699>) (deny|permit) (udp) ((A.B.C.D
A.B.C.D)|(any)|(host A.B.C.D)) [eq
(<1-65535>|rip|snmp|snmp-trap|tftp)] ((A.B.C.D A.B.C.D)
|(any)|(host A.B.C.D)) [eq (<1-65535>|rip|snmp|snmp-trap|tftp)]
no access-list (<100-199>|<2000-2699>) (deny|permit) (udp)
((A.B.C.D A.B.C.D)|(any)|(host A.B.C.D)) [eq
(<1-65535>|rip|snmp|snmp-trap|tftp)] ((A.B.C.D A.B.C.D)
|(any)|(host A.B.C.D)) [eq (<1-65535>|rip|snmp|snmp-trap|tftp)]
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

(<100-199>|<2000-2699>) : indicates the scope of the extended ACL.

(deny|permit) : ACL action, deny, permit.

udp: filter udp protocol message.

((A.B.C.D A.B.C.D)|(any)|(host A.B.C.D))…((A.B.C.D A.B.C.D)|(any)|(host A.B.C.D)): The former represents the source IP address and mask information, while the latter represents the destination IP address and mask information. The specific information is as follows.

- (A.B.C.D A.B.C.D): represents the source/destination IP address and mask. The mask adopts the anti-code mechanism. For example, 192.168.1.1 0.0.0.0, which means that only packets match 192.168.1.1 source/destination IP.
- host A.B.C.D: indicates that the source/destination IP address is A.B.C.D and the mask is 0.0.0.0.
- any: indicates that the source /destination IP address is 0.0.0.0 255.255.255.255, which means all IP addresses.

eq: specify equal to source/destination port.

(<1-65535>|rip|snmp|snmp-trap|tftp)：corresponding to different tcp message types:

- Port number(1-65535)，
- Routing Information Protocol (router, in.routed, 520)，
- Simple Network Management Protocol (161)，
- SNMP Traps (162)，
- Trivial File Transfer Protocol (69)。

【Description】

Configure extended ACL UDP protocol based on IPv4. For example, the configuration rule is as follows: access-list 101 deny udp host 192.168.1.1 eq tftp host 192.168.2.1 eq tftp, then when receiving the tcp tftp message from 192.168.1.1 to 192.168.2.1, the action executed is discarded. These rules only take effect when activated on a port using the ip-access-group command. And has the "first activation first effect" feature.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list 101 deny tcp host 192.168.1.1 eq tftp
host 192.168.2.1 eq tftp
```

26.3.6 Configure Character Type ACL Based on IPv4 Addresses

【Command】

When the protocol is gre, igmp, ipcomp, ospf, pim, rsvp and vrrp, the command format of the advanced access control list is:

access-list swos WORD **(deny|permit)** **(**<0-255>**|gre|igmp|ip|ipcomp|ospf|pim|rsvp|vrrp)** **(**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any)** **(**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any)** **[(fragments|(label** <1-65535>**)|log|(pkt-size** **((gt|lt)** <0-65535>**)|(rang** <0-65535> <0-65535>**))|precedence** <0-7>**|(tos (<**0-255>**|(rang** <0-255> <0-255>**)))]**

no access-list swos WORD **(deny|permit)** **(**<0-255>**|gre|igmp|ip|ipcomp|ospf|pim|rsvp|vrrp)** **(**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any)** **(**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any)** **[(fragments|(label** <1-65535>**)|log|(pkt-size** **((gt|lt)** <0-65535>**)|(rang** <0-65535> <0-65535>**))|precedence** <0-7>**|(tos (<**0-255>**|(rang** <0-255> <0-255>**)))]**

When the protocol is TCP, the command format of the advanced access control list is:

access-list swos WORD **(deny|permit)** **(**<0-255>**|tcp)** **((**A.B.C.D A.B.C.D)|A.B.C.D/M**|any)** **[((eq|gt|lt|ne)** <0-65535>**)|(rang** <0-65535> <0-65535>**)]** **(**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any)** **[((eq|gt|lt|ne)** <0-65535>**)|(rang** <0-65535> <0-65535>**)|established|fragments|(label** <1-65535>**)|log|(pkt-size** **((gt|lt)** <0-65535>**)|(rang** <0-65535> <0-65535>**))|precedence** **<0-7>|(tos** (<0-255>**|(rang** <0-255> <0-255>**)))]**

no access-list swos WORD **(deny|permit)** **(**<0-255>**|tcp)** **((**A.B.C.D A.B.C.D)|A.B.C.D/M**|any)** **[((eq|gt|lt|ne)** <0-65535>**)|(rang** <0-65535> <0-65535>**)]** **(**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any)** **[((eq|gt|lt|ne)** <0-65535>**)|(rang** <0-65535> <0-65535>**)|established|fragments|(label** <1-65535>**)|log|(pkt-size** **((gt|lt)** <0-65535>**)|(rang** <0-65535> <0-65535>**))|precedence** **<0-7>|(tos** (<0-255>**|(rang** <0-255> <0-255>**)))]**

When the protocol is UDP, the command format of the advanced access control list is:

access-list swos WORD **(deny|permit)** (<0-255>**|udp) (**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any) [((eq|gt|lt|ne)** <0-65535>**)|(rang** <0-65535> <0-65535>**)] ((A.B.C.D A.B.C.D)|A.B.C.D/M|any) [((eq|gt|lt|ne)** <0-65535>**)|(rang** <0-65535> <0-65535>**)|fragments|(label** <1-65535>**)|log|(pkt-size ((gt|lt)** <0-65535>**)|(rang** <0-65535> <0-65535>**))|precedence** <0-7>**|(tos** (<0-255>**|(rang** <0-255> <0-255>**))**)**]**

no access-list swos WORD **(deny|permit)** (<0-255>**|udp) (**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any) [((eq|gt|lt|ne)** <0-65535>**)|(rang** <0-65535> <0-65535>**)] ((A.B.C.D A.B.C.D)|A.B.C.D/M|any) [((eq|gt|lt|ne)** <0-65535>**)|(rang** <0-65535> <0-65535>**)|fragments|(label** <1-65535>**)|log|(pkt-size ((gt|lt)** <0-65535>**)|(rang** <0-65535> <0-65535>**))|precedence** <0-7>**|(tos** (<0-255>**|(rang** <0-255> <0-255>**))**)**]**

When the protocol is ICMP**, the command format of the advanced access control list is:**

access-list swos WORD **(deny|permit) (**<0-255>**|icmp) (**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any) (**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any) [fragments|(icmp-type** ICMP-TYPE**)|(label** <1-65535>**)|log|(pkt-size ((gt|lt)** <0-65535>**)|(rang** <0-65535> <0-65535>**))|precedence** <0-7>**|(tos (**<0-255>**|(rang** <0-255> <0-255>**))))]**

no access-list swos WORD **(deny|permit) (**<0-255>**|icmp) (**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any) (**(A.B.C.D A.B.C.D)|A.B.C.D/M**|any) [fragments|(icmp-type** ICMP-TYPE**)|(label** <1-65535>**)|log|(pkt-size ((gt|lt)** <0-65535>**)|(rang** <0-65535> <0-65535>**))|precedence** <0-7>**|(tos (**<0-255>**|(rang** <0-255> <0-255>**))))]**

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

Swos WORD: configure a character ACL.

(deny|permit) : ACL action, deny, permit.

(<0-255>|gre|icmp|igmp|ip|ipcomp|ospf|pim|rsvp|vrrp)：

- <0-255>：IANA assigned protocol number；
- gre：GRE packet；
- igmp：IGMP packet；
- ip：IP packet；

- ipcomp：IPComp packet；
- ospf：OSPF packet；
- pim：PIM packet；
- rsvp：RSVP packet；
- vrrp：VRRP packet。

(tcp)：TCP packet。

(udp)：UDP packet。

(icmp)：ICMP packet。

((A.B.C.D A.B.C.D)|A.B.C.D/M|any) ((A.B.C.D A.B.C.D)|A.B.C.D/M|any): The former represents the source IP address and mask information, while the latter represents the destination IP address and mask information. The specific information is as follows.

- A.B.C.D A.B.C.D: represents the source IP address and mask. The mask adopts the anti-code mechanism, such as 192.168.1.1 0.0.0.0 means only match 192.168.1.1 source IP message.
- host A.B.C.D: indicates that the source IP address is A.B.C.D and the mask is 0.0.0.0.
- any: indicates that the source IP address and mask are 0.0.0.0 255.255.255.255, which means all IP addresses.

[((eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)]…[((eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)： The former represents the source port number, while the latter represents the destination port number. The specific information is as follows.

- eq <0-65535>： specify equal to source/destination port.
- gt <0-65535>： specify larger than the source/destination port.
- lt <0-65535>: specify less than the source/destination port.
- ne <0-65535>:specify not equal to the source/destination port.
- rang <0-65535> <0-65535>: specify the range of source/destination ports. <0-65535> <0-65535> respectively indicate the beginning and end of the port range.

established: the message of TCP intermediate connection process. The type of SYN Flag in TCP header of ACL rule matching message is ack or rst.

icmp-type ICMP-TYPE: specifies the message type of ICMP message of ACL rule matching message, which is only valid if the message protocol is ICMP. If it is not configured, it means that any ICMP type message matches.

fragments: specifies whether this rule is only valid for non-first fragment messages. When this parameter is included, it means that this rule is only valid for non-first fragment message.

(label <1-65535 >): configure the priority.

log: specifies that the IP information of messages matched with this rule will be recorded in the log.

(pkt-size ((gt|lt) <0-65535>)|(rang <0-65535> <0-65535>)): matches the message length.

precedence <0-7 >: when ACL rules match messages, filter according to priority field.

(tos (<0-255>|(rang <0-255> <0-255>)): when an ACL rule matches a message, it is filtered according to the specified tos priority or priority range.

【Description】

The Access-list command is used to create name/character filter rule group. A group can support up to 32 rules. No is to delete a rule group. When the message matches the corresponding rule, the action will be executed. For example, the configuration rule is as follows: access-list swos AA deny ip 192.168.1.1 0.0.0.0 192.168.2.1 0.0.0.0. When the message from 192.168.1.1 is received and sent to 192.168.2.1, the action performed is discard. These rules only take effect when activated on a port using the ip-access-group command. And has the "first activation first effect" feature.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#access-list swos AA deny ip 192.168.1.1 0.0.0.0
192.168.2.1 0.0.0.0
```

26.3.7 Configure Character Type Standard ACL Based on Ipv6 Addresses

【Command】

```
ipv6 access-list NAME (deny|permit) (X:X::X:X/M|any)
no ipv6 access-list NAME (deny|permit) (X:X::X:X/M|any)
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

Ipv6 access-list NAME: configure an IPv6-based character standard ACL.

(deny|permit) : ACL action, deny, permit.

X:X::X:X/M: indicates the source IPv6 address and mask.

any: indicates any source IPv6 address.

【Description】

Access-list: the command is used to create character-type standard ACL filtering rule groups. Each group supports up to 32 rules. no means to delete a certain rule group. When the message matches the desired rule, the corresponding action will be executed. These rules only take effect when activated on a port using the ip-access-group command. And has the "first activation first effect" feature.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ipv6 access-list aa deny any
```

26.3.8 Configure Character Type Extended ACL Based on Ipv6 Addresses

【Command】

When the protocol is gre, igmp, ipcomp, ospf, pim, rsvp and vrrp, the command format of the advanced access control list is:

```
ipv6 access-list swos WORD (deny|permit)
(ip|gre|igmp|pim|rsvp|ospf|vrrp|ipcomp|<0-255>)
(X:X::X:X/M|(X:X::X:X X:X::X:X)|any) (X:X::X:X/M|(X:X::X:X
X:X::X:X)|any) [(label <1-65535>)|(precedence <0-7>)|(tos
(<0-255>|(range <0-255> <0-255>))|(pkt-size ((lt|gt)
<0-65535>)|(range <0-65535> <0-65535>))|fragments|log]
no ipv6 access-list swos WORD (deny|permit)
(ip|gre|igmp|pim|rsvp|ospf|vrrp|ipcomp|<0-255>)
(X:X::X:X/M|(X:X::X:X X:X::X:X)|any) (X:X::X:X/M|(X:X::X:X
X:X::X:X)|any) [(label <1-65535>)|(precedence <0-7>)|(tos
(<0-255>|(range <0-255> <0-255>))|(pkt-size ((lt|gt)
<0-65535>)|(range <0-65535> <0-65535>))|fragments|log]
```

When the protocol is ICMP**, the command format of the advanced access control list is:**

```
ipv6 access-list swos WORD (deny|permit) (<0-255>|icmp)
(X:X::X:X/M|(X:X::X:X X:X::X:X)|any) (X:X::X:X/M|(X:X::X:X
X:X::X:X)|any) [(icmp-type ICMP-TYPE)|(label
<1-65535>)|(precedence <0-7>)|(tos (<0-255>|(range <0-255>
```

<0-255>)|(pkt-size ((lt|gt) <0-65535>)|(range <0-65535> <0-65535>)|fragments|log]

no ipv6 access-list swos WORD (deny|permit) (<0-255>|icmp) (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) [(icmp-type ICMP-TYPE)|(label <1-65535>)|(precedence <0-7>)|(tos (<0-255>|(range <0-255> <0-255>)|(pkt-size ((lt|gt) <0-65535>)|(range <0-65535> <0-65535>)|fragments|log]

When the protocol is UDP**, the command format of the advanced access control list is:**

ipv6 access-list swos WORD (deny|permit) (<0-255>|udp) (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) [(eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)] (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) [(eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)] [(label <1-65535>)|(precedence <0-7>)|(tos (<0-255>|(range <0-255> <0-255>)|(pkt-size ((lt|gt) <0-65535>)|(range <0-65535> <0-65535>)|fragments|log]

no ipv6 access-list swos WORD (deny|permit) (<0-255>|udp) (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) [(eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)] (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) [(eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)] [(label <1-65535>)|(precedence <0-7>)|(tos (<0-255>|(range <0-255> <0-255>)|(pkt-size ((lt|gt) <0-65535>)|(range <0-65535> <0-65535>)|fragments|log]

When the protocol is TCP**, the command format of the advanced access control list is:**

ipv6 access-list swos WORD (deny|permit) (<0-255>|tcp) (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) [(eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)] (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) [(eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)] [established|(label <1-65535>)|(precedence <0-7>)|(tos (<0-255>|(range <0-255> <0-255>)|(pkt-size ((lt|gt) <0-65535>)|(range <0-65535> <0-65535>)|fragments|log]

no ipv6 access-list swos WORD (deny|permit) (<0-255>|tcp) (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) [(eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)] (X:X::X:X/M|(X:X::X:X X:X::X:X)|any) [(eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)] [established|(label <1-65535>)|(precedence <0-7>)|(tos (<0-255>|(range <0-255> <0-255>)|(pkt-size ((lt|gt) <0-65535>)|(range <0-65535> <0-65535>)|fragments|log]

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

Swos WORD: configure a character ACL.

(deny|permit) : ACL action, deny, permit.

(<0-255>|gre|icmp|igmp|ip|ipcomp|ospf|pim|rsvp|vrrp)：

- <0-255>：IANA assigned protocol number；
- gre：GRE packet；
- igmp：IGMP packet；
- ip：IP packet；
- ipcomp：IPComp packet；
- ospf：OSPF packet；
- pim：PIM packet；
- rsvp：RSVP packet；
- vrrp：VRRP packet。

(tcp)：TCP packet。

(udp)：UDP packet。

(icmp)：ICMP packet。

(X:X::X:X/M|(X:X::X:X X:X::X:X)|any)…(X:X::X:X/M|(X:X::X:X X:X::X:X)|any): The former represents the source IPv6 address and mask information, while the latter represents the destination IPv6 address and mask information. The specific information is as follows.

- X:X::X:X/M: indicates the source /destination IPv6 address and mask.
- X:X::X:X X:X::X:X: indicates the source /destination IPv6 address and mask.
- any: indicates any source /destination IPv6 address.

[((eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)]…[((eq|gt|lt|ne) <0-65535>)|(rang <0-65535> <0-65535>)： The former represents the source port number, while the latter represents the destination port number. The specific information is as follows.

- eq <0-65535>： specify equal to source/destination port.
- gt <0-65535>： specify larger than the source/destination port.
- lt <0-65535>: specify less than the source/destination port.
- ne <0-65535>:specify not equal to the source/destination port.
- rang <0-65535> <0-65535>: specify the range of source/destination ports. <0-65535> <0-65535> respectively indicate the beginning and end of the port range.

established: the message of TCP intermediate connection process. The type of SYN Flag in TCP header of ACL rule matching message is ack or rst.

icmp-type ICMP-TYPE: specifies the message type of ICMP message of ACL rule matching message, which is only valid if the message protocol is ICMP. If it is not configured, it means that any ICMP type message matches.

fragments: specifies whether this rule is only valid for non-first fragment messages. When this parameter is included, it means that this rule is only valid for non-first fragment message.

(label <1-65535 >): configure the priority.

log: specifies that the IP information of messages matched with this rule will be recorded in the log.

(pkt-size ((gt|lt) <0-65535>)|(rang <0-65535> <0-65535>)): matches the message length.

precedence <0-7 >: when ACL rules match messages, filter according to priority field.

(tos (<0-255>|(rang <0-255> <0-255>)): when an ACL rule matches a message, it is filtered according to the specified tos priority or priority range.

【Description】

The access-list command is used to create a name/character filter rule group. A group can support up to 32 rules. No is to delete a rule group. When the message matches the corresponding rule, the action will be executed. For example, the configuration rule is as follows: Ipv6 access-list swos qwer deny ip fe80::01 :: fe80::02 :: then when receiving the message from fe80::01 ::, to fe80::02 ::, the action executed is discard. These rules only take effect when activated on a port using the ip-access-group command. And has the "first activation first effect" feature.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#ipv6 access-list swos qwer deny ip fe80::01 :: fe80::02 ::
```

26.3.9 View All Configured ACL

【Command】

```
Show access-list
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

None

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#show access-list
```

26.3.10 Configure time-range

【Command】

Time-range NAME

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

NAME

【Description】

Time-range: means to configure a period of time during which the action of the ACL is executed, and when the time is up, the action of the ACL is not executed, which acts as a timing function. Support 20 groups of time-range, each group supports two types of time absolute (absolute time) and inquire (cycle time), and each group supports the creation of up to 16 time rules. Absolute time means to select a period of time, cycle time means to select a period of time, the weekly cycle time is divided into day, workday, non-work, supporting the configuration of time as well.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#time-range ac // create a time-range called ac
```

```
Switch(config-timerange_ac)#absolute start 12:00:00 1971-12-14
end 14:00:00 1971-12-15  // configure a period of time from 12:00:00
1971-12-14 to 14:00:00 1971-12-15.
Switch(config-timerange_ac)#periodic 09:00:00 to 12:00:00 daily
// configure a weekly cycle time from 09:00:00 to 12:00:00 every
morning
*Switch(config-timerange_ac)#exit
*Switch(config)#exit
*Switch#show time-range // view the current configuration of
time-range
Current time: 10:52 1970-01-01
time-range :ac
 periodic 09:00:00 to 12:00:00 daily
 absolute start 12:00:00 1971-12-14 end 14:00:00 1971-12-15
Switch#configure terminal
Switch(config)#time-range ac
Switch(config-timerange_ac)#no periodic 09:00:00 to 12:00:00  //
delete the sub-items of configuration cycle time within the
time-range
Switch(config-timerange_ac)#no absolute start 12:00:00 1971-12-14
end 14:00:00 1971-12-15  //Delete the configuration absolute time
subitem in the time-range
Switch(config-timerange_ac)#exit
Switch(config)#no time-range ac  //delete time-range ac
```

26.3.11 time-range Binds to the ACL

【Command】

access-list (<1-99>|<100-199>|<1300-1999>|<2000-2699>|NAME) **time-range** WORD

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<1-99>|<100-199>|<1300-1999>|<2000-2699>|NAME：ACL number.

WORD: Time range name.

【Description】

Bind a time-range to an ACL and perform the ACL action within the setting time. An ACL can only bind one time-range, and one time-range can bind multiple ACL. When deleted, if the time-range is referenced, the time-range is not allowed to be deleted and its subitems are allowed to be modified.

【Instance】

```
Switch(config)#access-list 10 time-range ad // bind the standard ACL numbered 10 to time-range ad.
Switch(config)#no access-list 10 time-range ad // unbind the standard ACL numbered 10 to time-range ad.
```

26.3.12 Activate IP ACL

【Command】

```
ip-access-group (<1-199>|<1300-2699>|NAME) in
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

(<1-199>|<1300-2699>) in: ACL rule group ID, in represents the ingress direction.

NAME：ACL number, SWOS name/character ACL.

【Description】

ACLS configured with time-range also need to be activated, meeting the rule of first activation first effect.

A port can only activate one IP address ACL and one MAC address ACL.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge8
Switch(config-ge8)#ip-access-group 1 in
```

26.3.13 Configure ACL based on MAC address

【Command】

mac access-list <3000-3699> **(deny|permit) (**(MAC MASK)**|any|(host** MAC**)) (**(MAC MASK)**|any|(host** MAC**)) [**<1536-65535>|NUM**]**

no mac access-list <3000-3699> **(deny|permit) (**(MAC MASK)**|any|(host** MAC**)) (**(MAC MASK)**|any|(host** MAC**)) [**<1536-65535>|NUM**]**

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

< 3000-3699 >: expand the label range of MAC ACL.

(deny|permit) : ACL action, deny, permit.

((MAC MASK)|any|(host MAC)) ((MAC MASK)|any|(host MAC)): the former indicates the source MAC address and mask information, while the latter indicates the destination MAC address and mask information, with details as follows.

- (MAC MASK): indicates the source/destination MAC address and mask of the message.
- any: indicates any source/destination MAC address.
- (host MAC): indicates the specified source/destination MAC address.

(<1-65535>|NUM|): Ethernet type. Hexadecimal/hexadecimal input.

【Description】

Configure a MAC address based ACL, and when the message matches the issued rule, the configured action is executed.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#MAC access-list 3001 deny any any 0x8100//Configure a message ACL that discards the MAC address of Ethernet type 0x8100
Switch(config)#no mac access-list 3001 deny any any 0x8100 //// delete a message ACL that discards a MAC address of Ethernet type 0x8100
```

26.3.14 View all configured MAC ACL

【Command】

show mac access-list [<3000-3699>]

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

< 3000-3699 >: layer 2 ACL number.

【Description】

Check all or specified MAC ACL information.

【Instance】

```
Switch> enable
Switch# show mac access-list
```

26.3.15 Time-range and MAC ACL binding

【Command】

mac access-list <3000-3699> **time-range** WORD

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

< 3000-3699 >: layer 2 ACL number.

WORD: Time range name.

【Description】

Bind a time-range to an MAC ACL and perform the MAC ACL action within the setting time. An MAC ACL can only bind one time-range, and one time-range can bind multiple MAC ACL. When deleted, if the time-range is referenced, the time-range is not allowed to be deleted and its subitems are allowed to be modified.

【Instance】

```
Switch(config)#mac access-list 3001 time-range ad // bind the standard MAC ACL numbered 3001 to time-range ad.
Switch(config)#no mac access-list 3001 time-range ad // unbind the standard MAC ACL numbered 3001 to time-range AD.
```

26.3.16 Activate MAC ACL

【Command】

```
mac-access-group <3000-3699> in
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

<3000-3699>: MAC ACL rule group ID.

in: indicates the direction of ingress.

【Description】

ACLS configured with time-range also need to be activated, meeting the rule of first activation first effect.

A port can activate only one ACL based on IP address and one ACL based on MAC address.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#mac-access-group 3001 in
```

26.3.17 View all Activated ACL

【Command】

```
show access-group
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

View the currently activated ACL port.

【Instance】

```
Switch> enable
Switch#show access-group
interface ge1
  ip-access-group sw in
  mac-access-group 3001 in
```

27 802.1X Authentication Configuration

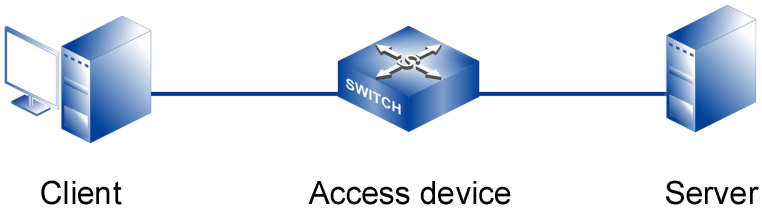
27.1 Overview

The IEEE802 LAN/WAN committee proposed the 802.1X protocol to address network security issues in wireless LANs. Later, the 802.1X protocol, as a common access control mechanism for LAN, was widely used in Ethernet, mainly to solve the authentication and security problems within Ethernet.

27.2 Principles

802.1X protocol is a port-based network access control protocol. "Port-based network access control" means that access to network resources is controlled by authenticating the accessed user device at the port level of LAN access equipment.

As shown in the following figure, 802.1X system is a typical Client/Server structure, including three entities: Client, Device and authentication Server.



- A client is an entity located at one end of a LAN segment, which is authenticated by a device at the other end of the link. A client is generally a user terminal device, and a user can initiate 802.1X authentication by starting the client software. Clients must support EAPOL (extensible authentication protocol over LAN) on local area network.
- The device end is another entity located at one end of the LAN segment, which

authenticates the connected client. The device end is usually a network device supporting 802.1X protocol, which provides a port for the client to access the local area network, which can be either a physical port or a logical port.

- Authentication server is an entity that provides authentication services for device. Authentication server, usually RADIUS server, is used to authenticate, authorize and charge users.

27.2.1 Basic Concepts of 802.1X

Controlled/Uncontrolled Ports

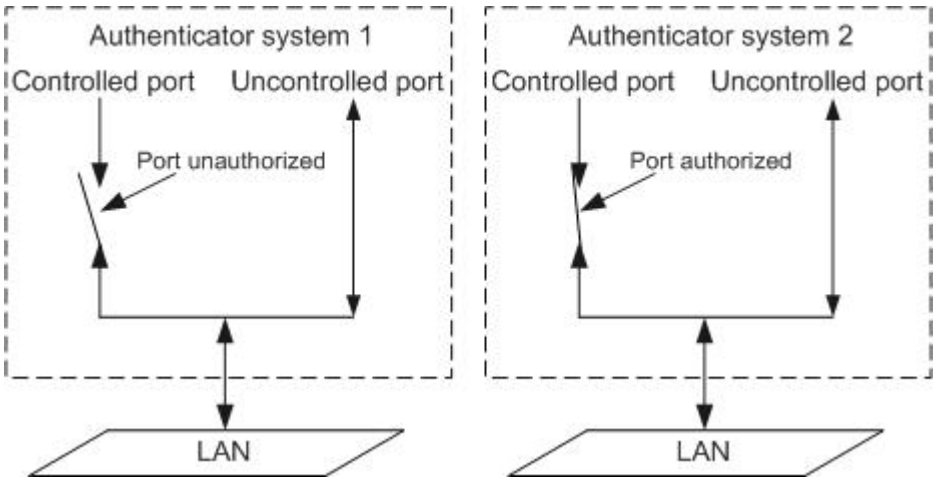
The device provides the client with a port to access the LAN, which is divided into two logical ports: controlled port and uncontrolled port.

- Uncontrolled ports are always in two-way communication state, which is mainly used to transmit EAPOL protocol frames to ensure that clients can always send or receive authentication messages.
- The controlled port is in a bidirectional communication state under the authorization state, and is used for transmitting service messages; It is forbidden to receive any message from the client in the unauthorized state.

Authorized/Unauthorized Status

The device end uses the authentication server to authenticate the client that needs to access the LAN, and controls the authorized/unauthorized state of the controlled port accordingly according to the authentication result (Accept or Reject).

The following figure shows the influence of different authorization states on messages passing through the controlled port. The figure compares the port status of two 802.1X authentication systems. The controlled port of system 1 is in unauthorized state (equivalent to opening the port switch), and the controlled port of system 2 is in authorized state (equivalent to closing the port switch).



27.2.2 Authentication Trigger Mode of 802.1X

The authentication process of 802.1X can be initiated by the client or the device. Authentication trigger methods supported by devices include the following two types:

- Client active trigger mode: the client actively sends EAPOL-Start message to the device to trigger authentication.
- Device active trigger mode: Device trigger mode is used to support clients that cannot actively send EAPOL-Start messages, such as the 802.1X client that comes with Windows system.

27.2.3 Authentication Method of 802.1X

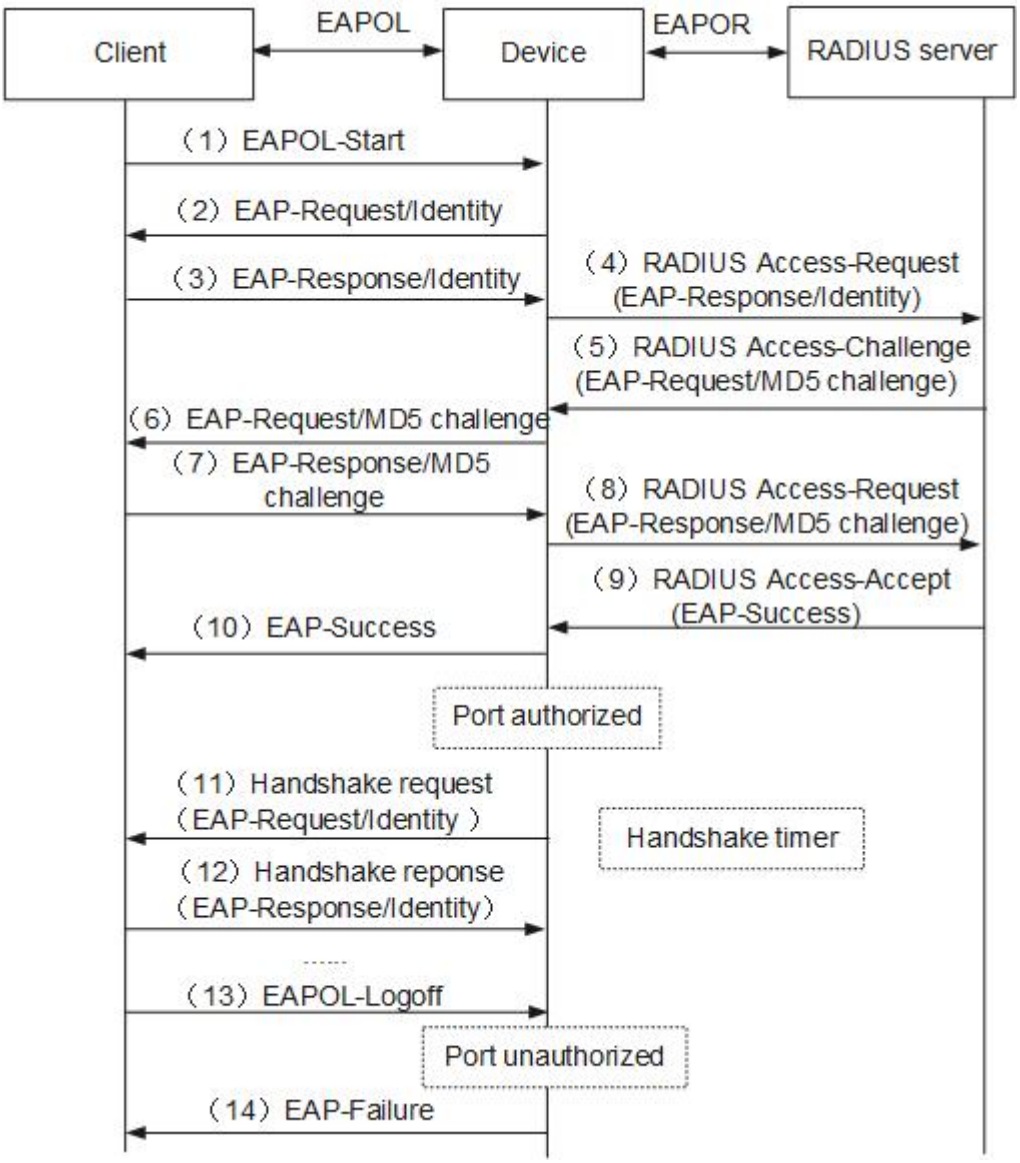
The 802.1X authentication system uses EAP (Extensible Authentication Protocol) to realize the exchange of authentication information among client, device and authentication server. The interaction form of EAP protocol messages among entities is as follows:

- Between the client and the device, EAP protocol messages use EAPOL encapsulation format and are directly carried in LAN environment.
- EAP protocol messages can interact in the following two ways between device and RADIUS server.
  - EAP relay: EAP protocol messages are relayed by the device, and the device carries EAP messages in RADIUS protocol using EAP POR (EAP over RADIUS) encapsulation format, and sends them to RADIUS server for authentication. The advantage of this authentication method is that the device is easy to handle and can support various EAP authentication methods, such as MD5-Challenge, EAP-TLS, PEAP, etc., but the server side is required to support the corresponding authentication methods.
  - EAP termination: EAP protocol messages are terminated by the device, and the device encapsulates the client authentication information in the standard RADIUS message, and authenticates with the server by means of PAP (Password Authentication Protocol) or CHAP (Challenge Handshake Authentication Protocol). The advantages of this authentication method are that the existing RADIUS servers can basically support PAP and CHAP authentication without upgrading the servers, but the device processing is complicated and cannot support other EAP authentication methods except MD5-Challenge.

The 802.1X system supports EAP relay and EAP termination to interact with the remote RADIUS server to complete authentication. The following figure describes the

process of the two authentication methods, taking the initiative of the client to initiate authentication as an example.

EAP Relay Authentication



The EAP relay authentication process is as follows:

5 When the user needs to access the external network, open the 802.1X client program, input the applied and registered user name and password, and initiate a connection request. At this time, the client program will send an authentication request frame (EAPOL-Start) to the device to start an authentication process.

6 After receiving the authentication request frame, the device will send out an Identity type request frame (EAP-Request/Identity) to request the user's client program to send the input user name.

7 In response to the request sent by the device, the client program sends the user

name information to the device through an Identity type response frame (EAP-Response/Identity).

8 The device end encapsulates the EAP message in the response frame sent by the client in a RADIUS access-request and sends it to the authentication server for processing.

9 After receiving the user name information forwarded by the device, the RADIUS server compares the information with the user name list in the database, finds the password information corresponding to the user name, encrypts the password with a randomly generated MD5 Challenge, and sends the MD5 Challenge to the device through RADIUS Access-Challenge message.

10 The device end forwards the MD5 Challenge sent by the RADIUS server to the client.

11 After receiving the MD5 Challenge from the device, the client uses the Challenge to encrypt the password part, generates EAP-Response/MD5 Challenge message, and sends it to the device.

12 The device end encapsulates this EAP-Response/MD5 Challenge message in a RADIUS access-request and sends it to the RADIUS server.

13 The RADIUS server compares the received encrypted password information with the local encrypted password information. if they are the same, the user is considered as a legitimate user, and sends a RADIUS Access-Accept message to the device.

14 After receiving the authentication pass message, the device sends EAP-Success frame to the client, and changes the port to an authorized state, allowing the user to access the network through the port.

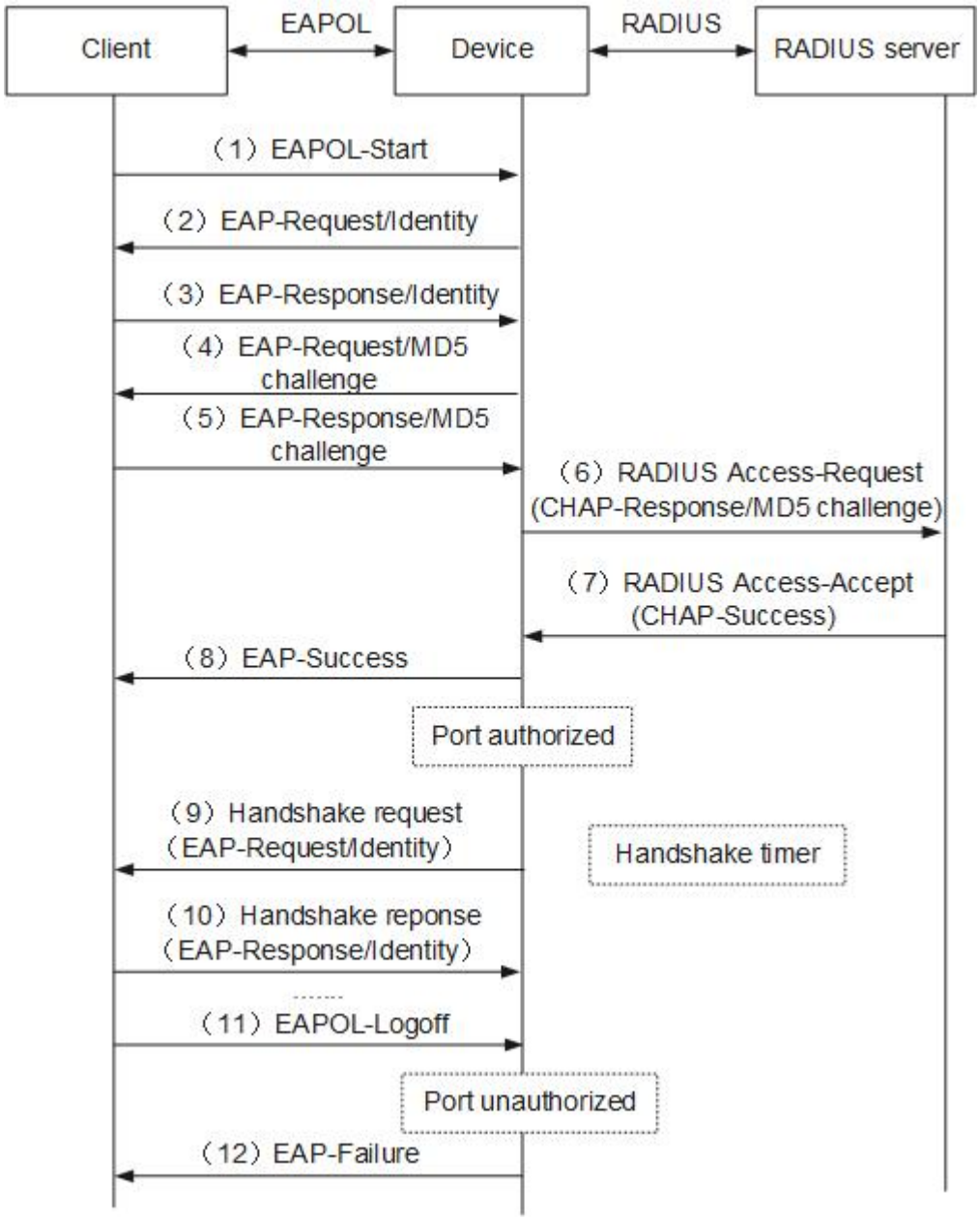
15 When the user is online, the device will monitor the online state of the user by sending handshake messages to the client regularly.

16 After receiving the handshake message, the client sends a response message to the device, indicating that the user is still online. By default, if the two handshake request messages sent by the device end are not answered by the client end, the device end will let the user go offline to prevent the user from going offline due to abnormal reasons and the device cannot perceive it.

17 Client can send EAPOL-Logoff frame to the device, and ask for offline actively.

18 The device changes the port status from authorized status to unauthorized status, and sends EAP-Failure message to the client.

EAP Termination Authentication



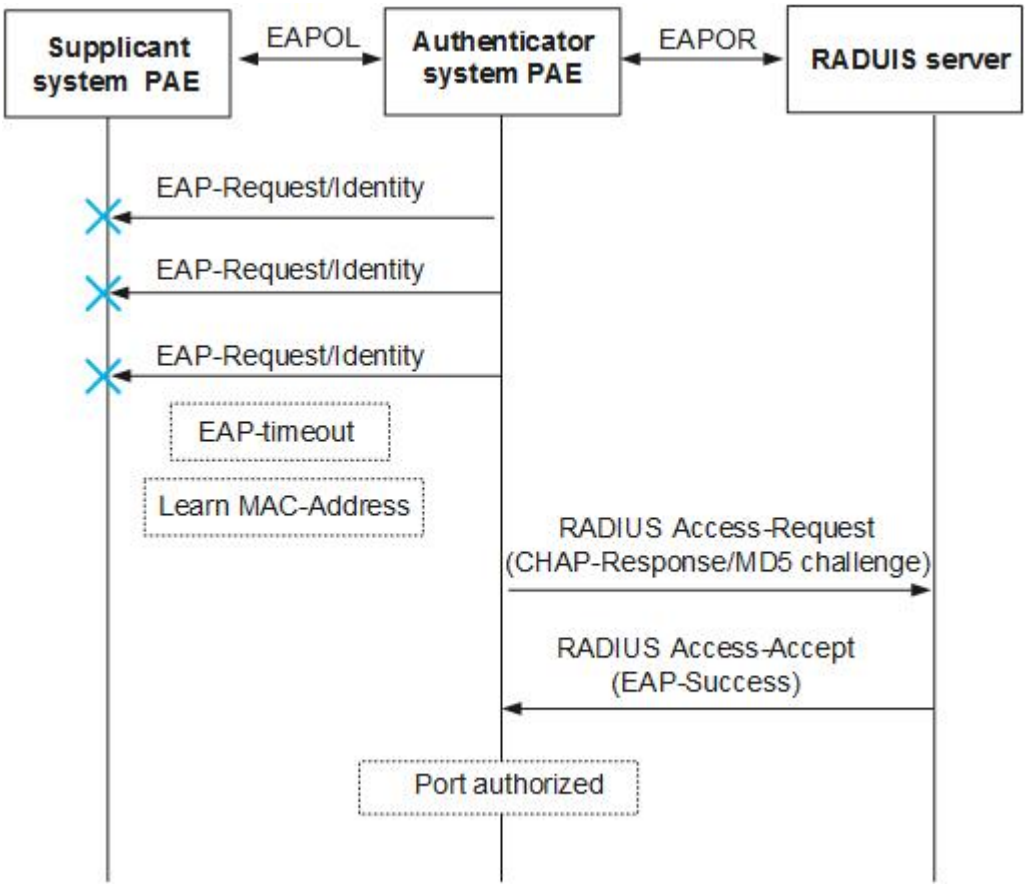
Compared with EAP relay authentication process, EAP termination method is different in that the MD5 challenge used to encrypt the user password information in step (4) is generated by the device, and then the device will send the user name, MD5 challenge and the encrypted password information of the client to the RADIUS server for related authentication processing.

27.2.4 MAC Bypass Authentication

MAC bypass authentication enables the terminals in 802.1X authentication system that cannot install and use 802.1X client software, such as printers, to authenticate with their own MAC address as user name and password.

In the process of 802.1X authentication, the device will first trigger the user to adopt 802.1X authentication mode, but if the user fails to perform 802.1X authentication for a long time, the MAC address of the user will be used as authentication information and sent to the authentication server as user name and password for authentication.

As shown in the figure below, MAC bypass authentication will be adopted when the device sends multiple authentication requests and the terminal does not respond.



27.2.5 802.1X Authentication Supports Dynamic VLAN Authorization

1、Guest VLAN

After the Guest VLAN function is enabled, when the user does not respond to the 802.1X authentication request, for example, if the client software is not installed, the device will add the port where the user is located to the Guest VLAN, so that the user can access the resources in the Guest VLAN, thus meeting the needs of unauthenticated users to obtain the client software, upgrade the client or execute other user upgrade programs.

2、Restrict VLAN

After the Restrict VLAN function is enabled, when the user authentication fails, such as entering an incorrect user name and password, the device will add the port where the user is located to the Restrict VLAN. The functions of Restrict VLAN and Guest VLAN are similar, which meet the requirement that users can access limited network resources before passing authentication. Generally, fewer network resources are deployed in the Restrict VLAN than in the Guest VLAN, thus restricting the access of unauthenticated users to network resources more strictly.

3、Critical VLAN

After the Critical VLAN function is enabled, when the authentication server does not respond, such as the network between the device and the authentication server is disconnected or the authentication server fails, the device will add the port where the user is located to the Critical VLAN and then access the resources in the Critical VLAN.

27.2.6 802.1X Rapid Deployment

NAC, as a network end-to-end access control scheme, improves the overall defense capability of the network. However, in the actual application process, the deployment workload of 802.1X clients is very heavy, which brings inconvenience to network construction. The 802.1X authentication rapid deployment function can solve the above problems, guide users to download and install clients by themselves, and realize the rapid deployment of clients. The 802.1X authentication rapid deployment function is realized through the following two functions:

- User restricted access
  Before 802.1X authentication succeeds, the end user can only access a specific IP address segment or a specific server through ACL, and provide services such as client download and upgrade or dynamic address allocation on the specific server.
- User HTTP access URL redirection function
  End users use IE to access the network before 802.1X authentication or after

authentication fails, and the device redirects the URL accessed by users to the configured URL (client download interface).

27.2.7 User Group Authorization Function

The device supports authorization control of users according to user groups, that is, after successful user authentication, the authentication server issues user groups and classifies users. Each user group can be associated with different ACL rules. Through the association between user groups and ACL rules, ACL authorization information control is realized for each type of users, that is, users of the same kind adopt the same authorization information.

27.3 Configure 802.1X Authentication

27.3.1 Global802.1X Authentication Enablement

【Command】

[no] dot1x system-auth-ctrl

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

dot1x system-auth-ctrl: command is used to enable global dot1x authentication function.

no dot1x system-auth-ctrl: command is used to disable global dot1x authentication function.

By default, global dotlx authentication function is disabled.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#dot1x system-auth-ctrl
```

27.3.2 802.1X Authentication Port Authorization Mode

【Command】

```
dot1x port-control (auto | force-authorized | force-unauthorized)
no dot1x port-control
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

auto: set port to enable 802.1x authentication mode and it is unauthorized mode by default.

force-authorized: set the port to forced authorized mode.

force-unauthorized: set the port to unauthorized forced mode

【Description】

dot1x port-control: command is used to set the access control mode of 802.1x on the specified port.

no dot1x port-control: the command is used to delete port 802.1x authentication function.

The port is not configured with 802.1x authentication by default.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x port-control auto
```

27.3.3 802.1X Authentication Port Controlled Direction

【Command】

```
dot1x port-control dir (both | in)
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

both: the controlled direction is bi-directional

in: the controlled direction is ingress.

【Description】

dot1x port-control dir: command is used to configure port controlled directions.

By default, the controlled direction of the port is ingress.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x port-control dir both
```

27.3.4 802.1X Authentication EAPOL Protocol Version

【Command】

```
dot1x protocol-version (1| 2)
no dot1x protocol-version
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

1: Configure EAPOL to 1

2: Configure EAPOL to 2

【Description】

dot1x protocol-version: command is used for the EAPOL protocol version of dot1x.

no dot1x protocol-version: command is used for the default EAPOL protocol version of dot1x.

By default, the EAPOL protocol message version is 2.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x protocol-version 2
```

27.3.5 802.1X Authentication Port Silent Time

【Command】

```
dot1x quiet-period <1-65535>
no dot1x quiet-period
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

1-65535: the silent time of the port, ranging from 1-65535 seconds, defaults to 60 seconds

【Description】

dot1x quiet-period: the command is used for the quiet period of the dot1x port.

no dot1x quiet-period: command is used for the default quiet time of the dot1x port.

By default, the silence time of the dot1x port is 60 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x quiet-period 120
```

27.3.6 802.1x Authorization Port Reauthentication Interval

【Command】

```
dot1x timeout re-authperiod <1-4294967295>
no dot1x timeout re-authperiod
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

1-4294967295: re-authentication interval for port, the range is 1-4294967295 seconds, the default value is 3600 seconds.

【Description】

dot1x timeout re-authperiod: command is used for re-authentication intervals on the dot1x port.

no dot1x timeout re-authperiod: command is used for the default re-authentication interval of dot1x port.

By default, the re-authentication interval on dot1x port is 3600 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x timeout re-authperiod 1200
```

27.3.7 802.1X Authorization Server Timeout Time

【Command】

```
dot1x timeout server-timeout <1-65535>
no dot1x timeout server-timeout
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

1-65535: the server timeout of the port, ranging from 1- 65535 seconds, defaults to 30 seconds.

【Description】

dot1x timeout server-timeout: The command is used for the server timeout on the dot1x port.

no dot1x timeout server-timeout: command is for the default server timeout of the dot1x port.

By default, the server timeout on dot1x port is 30 seconds.

【Instance】

```
Switch> enable
```

```
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x timeout server-timeout 60
```

27.3.8 802.1X Authorization Client Timeout Time

【Command】

dot1x timeout supp-timeout <*1-65535*>

no dot1x timeout supp-timeout

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

1-65535: the client timeout of the port, ranging from 1- 65535 seconds, defaults to 30 seconds

【Description】

dot1x timeout supp-timeout: command is used for the client timeout on the dot1x port.

no dot1x timeout supp-timeout: command is for the default clinet timeout of the dot1x port.

By default, the client timeout on dot1x port is 30 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x timeout supp-timeout 60
```

27.3.9 802.1X Authorization Message Retransmission Interval

【Command】

dot1x timeout tx-period <*1-65535*>

no dot1x timeout tx-period

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

1-65535: the retransmission interval of the port, the range is 1-65535 seconds, the default is 30 seconds

【Description】

dot1x timeout tx-period: command is used for retransmission intervals on the dot1x port.

no dot1x timeout tx-period: command is used for the default retransmission interval of the dot1x port.

By default, the retransmission interval on dot1x port is 30 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x timeout tx-period 60
```

27.3.10 802.1X Authorization Message Retransmission Interval

【Command】

dot1x reauthMax *<1-10>*

no dot1x reauthMax

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

1-65535: the number of retransmission of request/id message of the port, ranging from 1 to 10 seconds, it is 3 times by default

【Description】

dot1x reauthMax: command is used for the times of retransmissions of the dot1x port.

no dot1x reauthMax: command is uded for the default times of retransmissions

By default, the times of retransmissions on dot1x port is 3.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x reauthMax 4
```

27.3.11 802.1x Authorization Port Reauthentication Mode

【Command】

dot1x reauthentication

no dot1x reauthentication

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

dot1x reauthentication: command is used to enable re-authentication on the dot1x port.

no dot1x reauthentication: command is used to disable the re-authentication feature on the dot1x port.

By default, the re-authentication interval on dot1x port is 3600 seconds.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x reauthentication
```

27.3.12 802.1X Authentication Port Initialization

【Command】

dot1x initialize

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

dot1x initialize: command is used to initialize and unauthorize the dot1x port and attempt to re-authenticate on the dot1x port.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x initialize
```

27.3.13 802.1X Authorization Key Encryption Function

【Command】

dot1x keytxenabled (enable | disable)

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

dot1x keytxenabled enable: command is used to enable key encryption on the dot1x port (when clients interact with EPAOL messages).

dot1x keytxenabled disable: command is used to disable key encryption on the dot1x port.

Key encryption on the dot1x port is disabled by default.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#dot1x keytxenabled enable
```

27.3.14 Display 802.1X Authentication Global Information

【Command】

```
show dot1x
```

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show dot1x: command is used to display dot1x global information and radius client information.

【Instance】

```
Switch> enable
Switch#show dot1x
802.1X Port-Based Authentication Enabled
  RADIUS client address: not configured
```

27.3.15 Display 802.1X Authentication Detailed Information

【Command】

```
show dot1x all
```

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show dot1x all: command is used to display dot1x global information and radius client information, as well as port information.

【Instance】

```
Switch> enable
Switch#show dot1x all
802.1X Port-Based Authentication Enabled
  RADIUS client address: not configured
802.1X info for interface ge4
  portEnabled: true - portControl: Auto
  portStatus: Unauthorized - currentId: 90
  reAuthenticate: disabled
  reAuthPeriod: 3600
  abort:F fail:F start:F timeout:F success:F
  PAE: state: Connecting - portMode: Auto
  PAE: reAuthCount: 1 - rxRespId: 0
  PAE: quietPeriod: 60 - reauthMax: 2 - txPeriod: 30
  BE: state: Idle - reqCount: 0 - idFromServer: 0
  BE: suppTimeout: 30 - serverTimeout: 30
  CD: adminControlledDirections: in - operControlledDirections: in
  CD: bridgeDetected: false
  KR: rxKey: false
  KT: keyAvailable: false - keyTxEnabled: false
```

27.3.16 Display 802.1X Authentication Port Information

【Command】

show dot1x interface <IFNAME>

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

Ifname: specifies the port name

【Description】

show dot1x interface: command is used to display dot1x information for the specified port.

【Instance】

```
Switch> enable
Switch#show dot1x interface ge4
802.1X info for interface ge4
  portEnabled: true - portControl: Auto
  portStatus: Unauthorized - currentId: 92
  reAuthenticate: disabled
  reAuthPeriod: 3600
  abort:F fail:F start:F timeout:F success:F
  PAE: state: Connecting - portMode: Auto
  PAE: reAuthCount: 1 - rxRespId: 0
  PAE: quietPeriod: 60 - reauthMax: 2 - txPeriod: 30
  BE: state: Idle - reqCount: 0 - idFromServer: 0
  BE: suppTimeout: 30 - serverTimeout: 30
  CD: adminControlledDirections: in - operControlledDirections: in
  CD: bridgeDetected: false
  KR: rxKey: false
  KT: keyAvailable: false - keyTxEnabled: false
```

27.3.17 Display 802.1X Authentication Port Diagnosis Information

【Command】

show dot1x diagnostics interface <IFNAME>

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

Ifname:specifies the port name

【Description】

show dot1x diagnostics interface: command is used to display dot1x diagnostics information for the specified port.

【Instance】

```
Switch> enable
Switch#show dot1x diagnostics interface ge4
802.1X Diagnostics for interface ge4
 authEnterConnecting: 707
 authEaplogoffWhileConnecting: 355
 authEnterAuthenticating: 0
 authSuccessWhileAuthenticating: 0
 authTimeoutWhileAuthenticating: 0
 authFailWhileAuthenticating: 0
 authEapstartWhileAuthenticating: 0
 authEaplogoggWhileAuthenticating: 0
 authReauthsWhileAuthenticated: 0
 authEapstartWhileAuthenticated: 0
 authEaplogoffWhileAuthenticated: 0
 BackendResponses: 0
 BackendAccessChallenges: 0
 BackendOtherrequestToSupplicant: 0
 BackendAuthSuccess: 0
 BackendAuthFails: 0
```

27.3.18 Display 802.1X Authentication Port Session Information

【Command】

show dot1x sessionstatistics interface <IFNAME>

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

Ifname:specifies the port name

【Description】

show dot1x sessionstatistics interface: command is used to display dot1x sessionstatistics information for the specified port.

【Instance】

```
Switch> enable
```

```
Switch#show dot1x sessionstatistics interface ge4
802.1X session statistics for interface ge4
session authentication method: Local server
session time: 0 secs
session user name:
 session terminate cause: Port failure
```

27.3.19 Display 802.1X Authentication Port Message Statistics

【Command】

show dot1x statistics interface <IFNAME>

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

Ifname:specifies the port name

【Description】

show dot1x statistics interface: command is used to display dot1x message for the specified port.

【Instance】

```
Switch> enable
Switch#show dot1x statistics interface ge4
802.1X statistics for interface ge4
  EAPOL Frames Rx: 0 - EAPOL Frames Tx: 0
  EAPOL Start Frames Rx: 0 - EAPOL Logoff Frames Rx: 0
  EAP Rsp/Id Frames Rx: 0 - EAP Response Frames Rx: 0
  EAP Req/Id Frames Tx: 719 - EAP Request Frames Tx: 0
  Invalid EAPOL Frames Rx: 0 - EAP Length Error Frames Rx: 0
  EAPOL Last Frame Version Rx: 0 - EAPOL Last Frame Src:
0000.0000.0000
```

27.3.20 RADIUS Server Regeneration Interval

【Command】

```
radius-server deadtime <0-1440>
no radius-server deadtime
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

0-1440: the regeneration interval of RADIUS, ranging from 0-1440 minutes, and the default value is 0

【Description】

radius-server deadtime: command is used to configure the interval between the radius unreachable is restored to reachable.

no radius-server deadtime: command is used to delete the interval.

By default, the regeneration interval is 0.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#radius-server deadtime 5
```

27.3.21 RADIUS Server

【Command】

```
radius-server host <HOSTNAME> {key STRING | auth-port PORTNO |
timeout SEC | retransmit RETRIES}
no radius-server host <HOSTNAME> (auth-port PORTNO | )
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

hostname: IP address or host name of RADIUS server

STRING: Shared key with RADIUS server

PORTNO: UDP port of RADIUS authentication, the value range is 0-65535

SEC: timeout interval of RADIUS server, value range is 1-1000, the default value is 5 seconds

RETRIES: the number of times the RADIUS server retransmits over the timeout，the value range is 1-100, which is 3 times by default

【Description】

radius-server host: command is used to configure the RADIUS server.

no radius-server host: command is used to configure the RADIUS server.

By default, theRADIUS server is not configured.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#radius-server host 192.168.1.100 key 123456
auth-port 1812
```

28 Alarm Configuration

28.1 Enable Port Alarm

【Command】

system alarm enable

【View】

ge (Gigabit Ethernet) port view

xe (10 Gigabit Ethernet) port view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

Enabling port up/down the alarm light on the panel will be lit immediately when the port down occurs. When the port up occurs, the alarm light on the panel will be turned off. By default, it is turned off. However, the warning light will also be turned on when the power alarm is enabled, so it is recommended not to turn them on at the same.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge 1
Switch(config-ge1)#system alarm enable
```

28.2 Disable Port Alarm

【Command】

system alarm disable

【View】

ge (Gigabit Ethernet) port view

xe (10 Gigabit Ethernet) port view

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

Enabling port up/down the alarm light on the panel will be lit immediately when the port down occurs. When the port up occurs, the alarm light on the panel will be turned off. By default, it is turned off. However, the warning light will also be turned on when the power alarm is enabled, so it is recommended not to turn them on at the same.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge 1
Switch(config-ge1)#system alarm disable
```

28.3 Enable Power Alarm

【Command】

power <1-2> alarm enable

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<1-2>: means power supply 1, 2

【Description】

Enabling power<1- 2>alarm, light on the panel will be lit immediately when the port is down. When power is specified to up, the alarm light on the panel will be turned off. By default, it is turned off. However, the warning light will also be turned on when the port alarm is enabled, so it is recommended not to turn them on at the same.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#power 1 alarm enable
```

28.4 Power off Warning

【Command】

```
power <1-2> alarm disable
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<1-2>: means power supply 1, 2

【Description】

Enabling power<1- 2>alarm, light on the panel will be lit immediately when the port is down. When power is specified to up, the alarm light on the panel will be turned off. By default, it is turned off. However, the warning light will also be turned on when the port alarm is enabled, so it is recommended not to turn them on at the same.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#power 1 alarm disable
```

29 IPDT Configuration

29.1 Overview

IPDT (IP-Detection) specifies the detection target address and sends icmp echo request message to the target address regularly. If the icmp echo reply message is not received within the scheduled time, it will notify the linked application module; Then it continues to send icmp echo request. If the icmp echo reply message is received, it will notify the linked application module. The address detection function is mainly used to link with vrrp to ensure that when the uplink next hop device of the master is abnormal, it can be quickly switched to backup. This is a typical requirement for detecting abnormal vrrp uplink linkage. When the uplink of the vrrp group fails, the master cannot perceive it. At this time, auxiliary functions are required to detect exceptions and notify the VRRP group.

29.2 Configure IPDT

29.2.1 Create Session and Enter the Session View

【Command】

```
ip-detection <1-8>
no ip-detection <1-8>
```

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<1-8>: IPDT session ID, value range 1-8.

【Description】

ip-detection <1-8>: Create IPDT session and enter the session view.

no ip-detection <1-8>: delete specified IPDT session.

【Instance】

Create IPDT session 1.

```
*Switch#configure terminal
*Switch(config)#ip-detection 1
*Switch(config-ip-detection)#
```

29.2.2 Configure Source IP

【Command】

source-ip A.B.C.D

no source-ip

【View】

IPDT configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: The source IP address that sends probe packet.

【Description】

source-ip A.B.C.D: configure the source IP address for sending ICMP request message.

no source-ip: delete the source IP address of the IPDT session.

【Instance】

Configure the source IP address of IPDT Session 1 to 192.168.1.253.

```
*Switch#configure terminal
*Switch(config)#ip-detection 1
*Switch(config-ip-detection)#source-ip 192.168.1.253
```

29.2.3 Configure Destination IP

【Command】

target-ip A.B.C.D

no target-ip

【View】

IPDT configuration view

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: Destination IP address of the probe packet.

【Description】

target-ip A.B.C.D: configure the destination IP address of ICMP request message.

no target-ip: delete the destination IP address of the IPDT session.

【Instance】

Configure the destination IP address of IPDT Session 1 to 192.168.1.161.

```
*Switch#configure terminal
*Switch(config)#ip-detection 1
*Switch(config-ip-detection)#source-ip 192.168.1.161
```

29.2.4 Configure the Number of Per Detection

【Command】

echo-time <1-3>

no echo-time

【View】

IPDT configuration view

【Default Level】

2: Configuration level

【Parameter】

<1-3>: the number of times sent for each detection, with a value range of 1-3.

【Description】

echo-time <1-3>: configure the number of ICMP request messages sent for each detection.

no echo-time: restore the default number of ICMP request messages sent for each detection, which is 1 by default.

【Instance】

Set the number of detections sent by IPDT Session 1 each time to 1.

```
*Switch#configure terminal
*Switch(config)#ip-detection 1
*Switch(config-ip-detection)# echo-time 1
```

29.2.5 Configure Time Interval

【Command】

```
echo-interval <5-15>
no echo-interval
```

【View】

IPDT configuration view

【Default Level】

2: Configuration level

【Parameter】

<5-15>: interval time of detection packet, unit: 100ms, value range: 5-15.

【Description】

echo-interval <5-15>: configure the time interval of ICMP request messages sent for each detection.

no echo-interval: restore the default time interval for sending ICMP request message each time. The default value is 10.

【Instance】

Set the time interval of detections sent by IPDT Session 1 each time to 1 second.

```
*Switch#configure terminal
*Switch(config)#ip-detection 1
*Switch(config-ip-detection)# echo-interval 10
```

29.2.6 Enable/Disable IPDT Session

【Command】

enable

disable

【View】

IPDT configuration view

【Default Level】

2: Configuration level

【Parameter】

No.

【Description】

enable: :enable this IPDT session.

disable: disable this IPDT session.

【Instance】

Enable IPDT Session 1.

```
*Switch#configure terminal
*Switch(config)#ip-detection 1
*Switch(config-ip-detection)# enable
```

29.2.7 Display IPDT Session

【Command】

show ip-detection [<1-8>]

【View】

Privileged user mode

【Default Level】

1: view level

【Parameter】

<1-8>: The value range of IPDT ID is 1-8.

【Description】

show ip-detection <1-8>: view specified IPDT session.

show ip-detection: view all IPDT sessions.

【Instance】

View IPDT sessions.

```
*Switch#show ip-detection
Session <1>
 status is Enable
 source address is 192.168.1.253
 target address is 192.168.1.161
 echo time is 1
 echo interval is 1000 msec
 target is UP
 notify:     0
 echo:       12
 reply:      12
 send failed: 0
 not to me:   0
```

30 RMON Configuration

30.1 Overview

RMON is based on SNMP architecture and shares a set of network management workstations (NMS) with SNMP to remotely manage devices.

SNMP is the most widely used network management protocol in the Internet, which collects and counts network communication information through agent software embedded in devices. The management software sends a query signal to the MIB of the agent through polling to obtain the information, and realizes the management of the network through the obtained information. Although MIB counter records the sum of statistical data, it can't analyze the daily communication historically. In order to comprehensively check the traffic and the change of traffic in a day, the network management software needs constant polling to analyze the network status through the obtained information.

There are two obvious disadvantages in polling with SNMP:

- It takes up a lot of network resources. In a large network, a large number of network communication messages will be generated by polling, which will lead to network congestion and even network blockage. Therefore, SNMP is not suitable for managing large networks and recovering a large amount of data, such as routing table information.
- The burden of managers is increased, and the task of collecting data in SNMP polling is completed by network managers through network management software. If network managers monitor more than three network segments, it may happen that network managers cannot complete the task because of the heavy burden.

In order to improve the availability of management information, reduce the burden of management stations, and meet the needs of network administrators to monitor the performance of multiple network segments, IETF developed RMON to solve the limitations of SNMP in the ever-expanding distributed interconnection, and mainly

realized the function of monitoring the data traffic of one network segment and even the whole network. The following are the RMON characteristics:

- SNMP is the basis of RMON implementation, and RMON is an enhancement of SNMP function.
  RMON is implemented based on SNMP architecture and compatible with existing SNMP framework. It is still composed of NMS, a network management workstation, and Agent agents running on various network devices. Because RMON does not use another mechanism, NMS and SNMP of network management workstation are shared, and network managers do not need to learn extra, so the implementation is relatively simple.
- RMON enables SNMP to monitor remote network devices more effectively and actively, providing an efficient means to monitor the operation of networks.
  RMON protocol stipulates that the managed device can automatically send Trap information when the alarm threshold is reached, so the management device does not need to obtain MIB variable values through polling for comparison many times, thus reducing the communication traffic between the management device and the managed device, and achieving the purpose of simple and effective management of large-scale interconnection networks.

RMON allows multiple monitors, and monitors can collect data in the following two ways:

- With a special RMON Probe, NMS can directly obtain management information and control network resources from RMON Probe, which can obtain all information of RMON MIB.
- Embedding RMON Agent directly into network devices makes them become network devices with RMON Probe function. NMS uses SNMP to exchange data information and collect network management information. This method is limited by device resources, so it is generally impossible to obtain all the data of RMON MIB, and basically only collect the information of four groups (alarm, event, history and statistics).

The device adopts the second method, and realizes the RMON Agent function on the device. With this function, the management device can obtain information such as overall traffic, error statistics and performance statistics on the network segment connected with the managed network device interface, and then realize network monitoring.

30.2 Principles

Before configuring RMON, it is necessary to understand the basic concepts of statistics, history, alarms and events defined by RMON specification.

RMON mainly implements statistics and alarm functions, which are used for remote monitoring and management of managed devices by management devices in the network.

The statistical function of RMON can be realized by RMON statistical group or RMON historical group, which can be divided into Ethernet statistical function and historical statistical function.

- Ethernet statistics function (corresponding to the statistics group in RMON MIB): the system counts the basic statistical information of each monitored network. The system will continuously count the traffic of a certain network segment and the distribution of various types of packets, or the number of error frames and collisions of various types. The statistical objects include the number of network collisions, CRC error messages, too small (or too large) data messages, broadcast and multicast messages, and the number of bytes and messages received.
- Historical statistics function (corresponding to the historical group in RMON MIB): The system samples and collects network status statistics regularly and stores them for subsequent processing. The system will make statistics on various traffic information periodically, including bandwidth utilization, error packet number and total package number.

RMON alarm function includes event definition function and alarm threshold setting function, which is realized by the combination of these two sub-functions.

- Event definition function (corresponding to event group in RMON MIB): event group controls events and prompts from devices and provides all events generated by RMON Agent. When an event occurs, you can record a log or send a Trap to the network management station.
- Set alarm threshold function (corresponding to alarm group in RMON MIB): the system monitors the specified alarm variable (OID corresponding to any alarm object). After the user defines a set of thresholds and sampling time of the specified alarm in advance, the system will obtain the value of the specified alarm variable according to the defined time period, and trigger an upper limit alarm event when the value of the alarm variable is greater than or equal to the upper limit threshold; When the value of the alarm variable is less than or equal to the lower limit threshold, a lower limit alarm event is triggered. RMON Agent will record the monitored state as log or send Trap to network management station.

Multiple RMON groups are defined in the RMON specification (RFC2819), and devices implement the statistics, history, alarm, and event groups supported in the public MIB. These groups are introduced separately below.

- Statistical

  The statistics group stipulates that the system will continuously make statistics on various traffic information of the Ethernet interface, and store the statistical results in the Ethernet statistics table for the management device to check at any

time. The statistical information includes the number of network conflicts, the number of CRC error messages, the number of data messages that are too small (or too large), the number of broadcast and multicast messages, and the number of received bytes and received messages, etc.

After the statistics table item is successfully created under the specified interface, the statistics group counts the number of messages for the current interface, and it counts the result as a continuous cumulative value.

- History

  The history group regularly collects and stores network status statistics for subsequent processing.

  The history group contains two tables:

  - historyControlTable: it is mainly used to set control information such as sampling interval time.
  - ethernetHistoricalTable: it is mainly used to store the network status statistics collected regularly by etherHistoryTable group, and provide the network administrator with historical data about other statistical information such as network segment traffic, error packets, broadcast packets, utilization rate and collision times.

- Event

  Events defined by event groups are used in alarm group configuration items and extended alarm group configuration items. When the monitored objects meet the alarm conditions, the events will be triggered. RMON event management is to add events in the specified row of the event table and define how to handle the events:

  - **log**: send only logs
  - **trap**: only send trap message to NMS
  - **log-trap**: send both log and trap message to NMS
  - none: do not do anything

- Alarm

  The alarm group allows a predefined set of thresholds for alarm variables (which can be any object of local MIB) to be monitored. After the user defines AlarmTable (AlarmTable), the system will obtain the monitored value of the alarm variable according to the defined time period. When the value of the alarm variable is greater than or equal to the upper limit threshold, an upper limit alarm event will be triggered. When the value of the alarm variable is less than or equal to the lower limit, a lower limit alarm event is triggered, and the alarm management will process the event accordingly according to the definition of the event.

- Extended alarm group
  Based on RFC2819, the extended alarm group adds the function of setting alarm objects and alarm lifetime with expressions. It includes a prialarmTable, which has the following items more than the alarm table:
  - The expression string of extended alarm variables can be four expressions (+, -,*,/and parentheses) composed of several simple alarm variables OIDs.
  - Expand the description string of the alarm line.
  - Sampling type is rate of change.
  - The alarm state types are extended, including two types: forever and cycle. For cycle type, when the specified time of extended alarm state cycle has passed, no alarm will be generated and this table line will be deleted.

30.3 Configure RMON

30.3.1 RMON Alarm Group

【Command】

rmon alarm <INDEX> <ALARM-VARIABLE> interval <SECONDS> **{absolute | delta} rising-threshold <RISING_THRES> event** <EVENT_INDEX> **falling-threshold** <FALL_THRES> **event** <EVENT_INDEX> **(owner** <NAME>**|)**

no rmon alarm <INDEX>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

Index: warning group index, the range is 1-65535.

ALARM-VARIABLE: the format is etherStatsEntry.n.n

SECONDS: sampling time interval ranging from 1-4294967295.

Delta: sampling type is variable value (the current sample value of the selected variable relative to the last sample value)

absolute: the sampling type is absolute.

rising-threshold RISING_THRES: upper threshold, the value range is 0 ~ 2147483647.

EVENT_INDEX: index of event groups corresponding to the upper threshold, the range is 1-65535.

falling -threshold FALL_THRES: lower threshold, the value range is 0 ~ 2147483647.

EVENT_INDEX: index of event groups corresponding to the lower threshold, the range is 1-65535.

NAME: character string, creator of the row

【Description】

`rmon alarm`: command is used to configure alarm group.

`no rmon alarm`: command is used to delete alarm group.

By default alarm group is not configured.

The alarm variable format support string format (not OID format), formats are etherStatsEntryr.instance.integer or etherStatsString.instance, integer the range is 1-21, corresponding to the etherStatsString below respectively.

etherStatsString supports the following:

"etherStatsIndex"

"etherStatsDataSource"

"etherStatsDropEvents"

"etherStatsOctets"

"etherStatsPkts"

"etherStatsBroadcastPkts"

"etherStatsMulticastPkts"

"etherStatsCRCAlignErrors"

"etherStatsUndersizePkts"

"etherStatsOversizePkts"

"etherStatsFragments"

"etherStatsJabbers"

"etherStatsCollisions"

"etherStatsPkts64Octets"

"etherStatsPkts65to127Octets"

"etherStatsPkts128to255Octets"

"etherStatsPkts256to511Octets"

"etherStatsPkts512to1023Octets"

"etherStatsPkts1024to1518Octets"

"etherStatsOwner"

"etherStatsStatus"

instance is the index of interface RMON statistics group.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#rmon alarm 1 etherStatsIndex.1 interval 20 delta
rising-threshold 200 event 1 falling-threshold 20 event 1
```

30.3.2 RMON Statistical Group

【Command】

```
rmon collection stats <INDEX> {owner <NAME> | }
no rmon collection stats <INDEX>
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

INDEX:statistics group Index, the range is 1-65535.

NAME: character string, creator of the row

【Description】

rmon collection stats: command is used to configure the port statistics group.

no rmon collection stats: command is used to cancel the port statistics group.

The port is not configured with statistics group by default.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#rmon collection stats 1
```

30.3.3 RMON History Group

【Command】

```
rmon collection history <INDEX> {buckets <NUMBER> | interval
<SECONDS> | owner <NAME> | }
no rmon collection history <INDEX>
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

INDEX: statistics group index, the range is 1-65535.

NUMBER: set the historical table capacity corresponding to the history group, ranging from 1-65535.

SECONDS: sets the historical group statistical cycle value in the range of 1-3600 seconds.

NAME: creator of the row

【Description】

rmon collection history: command is used to configure the port history group.

no rmon collection stats: command is used to delete the port statistics group.

The port is not configured with history group by default.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#interface ge1
Switch(config-ge1)#rmon collection history 1 buckets 10 interval
60
```

30.3.4 RMON Event Group

【Command】

```
rmon event <INDEX> {description <STRING> | log | trap <COMMUNITY>}
(owner <NAME> | )
no rmon event <INDEX>
```

【View】

Ethernet port configuration mode

【Default Level】

2: Configuration level

【Parameter】

INDEX：event group index, the range is 1-65535.

Log: log events. When events are triggered, the system logs them.

STRING: event description.

community: Trap event. When the event is triggered, the system will send it with community as the group name

NAME: creator of the row

【Description】

rmon event: command is used to configure the event group.

no rmon event: command is used to cancel the event group.

The port is not configured with event group by default.

【Instance】

```
Switch> enable
Switch#configure terminal
Switch(config)#rmon event 2 log
```

30.3.5 Display RMON Alarm Group Information

【Command】

show rmon alarm

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show rmon alarm: command is used to display alarm group information.

【Instance】

```
Switch> enable
Switch#show rmon alarm
 alarm Index = 1
alarm status = VALID
       alarm Interval = 20
       alarm Type is Delta
       alarm Value  = 0
       alarm Rising Threshold = 200
       alarm Rising Event = 1
       alarm Falling Threshold = 20
       alarm Falling Event = 1
       alarm Owner is RMON_SNMP

 alarm Index = 2
```

```
alarm status = VALID
       alarm Interval = 20
       alarm Type is Delta
       alarm Value  = 0
       alarm Rising Threshold = 200
       alarm Rising Event = 1
       alarm Falling Threshold = 20
       alarm Falling Event = 1
       alarm Owner is RMON_SNMP
```

30.3.6 Display RMON Statistics Information

【Command】

show rmon statistics

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show rmon statistics: command is used to display statistics group information.

【Instance】

```
Switch> enable
Switch#show rmon statistics
   rmon collection index 1
   stats->ifindex = 5002
   input packets 00, bytes 00, dropped 00, multicast packets 00
   output packets 00, bytes 3406566434058944, multicast packets
00 broadcast packets 00
```

30.3.7 Display RMON History Group Information

【Command】

show rmon history

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show rmon history: command is used to display history group information.

【Instance】

```
Switch> enable
Switch#show rmon history
       history index = 1
       data source ifindex = 5002
       buckets requested = 50
       buckets granted = 50
       Interval = 1800
       Owner RMON_SNMP
```

30.3.8 Display RMON Event Group Information

【Command】

show rmon event

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

None

【Description】

show rmon event: command is used to display event group information.

【Instance】

```
Switch> enable
Switch#show rmon event
 event Index = 1
      Description RMON_SNMP
       Event type Log
      Last Time Sent = 07:43:20
      Owner  RMON_SNMP
```

31 Log Configuration

31.1 Log File Size Limit

【Command】

log file size <10-10000>

no log file size

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

<10-10000>：log file size, the unit is KB.

【Description】

log file size: the command is used to set the maximum size of logfile in KB.

no log file size: the command is used to delete the setting of logfile size and restore it to the default size, namely 2M.

【Instance】

```
# Configure the logfile size to 5M
Switch> enable
Switch#configure terminal
Switch(config)#log file size 5000

#Delete settings of logfile size
Switch> enable
Switch#configure terminal
Switch(config)#no log file size
```

31.2 Log stdout Display

【Command】

log stdout

no log stdout

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

No.

【Description】

log stdout: the command is used to open the switch and display the log information in stdout.

no log stdout: the command is used to close the switch and not to display the log information in stdout.

【Instance】

```
# Configure to open the log information displayed in stdout
Switch> enable
Switch#configure terminal
Switch(config)#log stdout

# close the log information displayed in stdout
Switch> enable
Switch#configure terminal
Switch(config)#no log stdout
```

31.3 LogInformation Highest Display Level

【Command】

Log trap (alerts | critical | debugging | emergencies | errors | informational | notifications | warnings)

no log trap

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

No.

【Description】

log trap: the command sets the maximum display level of log information.

no log trap: the command is used to remove the settings for the highest display level of log information and restore it to the default level, which is the "debug" level.

【Instance】

```
# set the log information for the highest level "informational"
Switch> enable
Switch#configure terminal
Switch(config)#log trap informational

# Delete the highest level Settings for log information
Switch> enable
Switch#configure terminal
Switch(config)#no log trap
```

31.4 Log Level Record Display

【Command】

log record-priority
no log record-priority

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

No.

【Description】

log record-priority: the command is used to open the level of log information, that is, to display the information according to the level of log information.

no log record-priority: the command is used to close the level of log information, and all log information is unified as debug level.

【Instance】

```
# Configure to open log information record-priority
Switch> enable
Switch#configure terminal
Switch(config)#log record-priority

# Close log information record-priority
Switch> enable
Switch#configure terminal
Switch(config)#no log record-priority
```

31.5 Syslog Server Download Log

【Command】

log syslog server <A.B.C.D> [<PORT>]
no log syslog server

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D：syslog server IP address

PORT: The port used by the syslog server.

【Description】

log syslog server: the command is used to set the IP address of the remote syslog server. After executing the command, the log information of the system will be sent to the syslog server with the specified IP address for processing remotely. The parameter A.B.C.D specifies the IP address used by the syslog server, and the parameter PORT specifies the port used by the syslog server.

no log syslog server: the command is used to delete the configuration of the remote syslog server. After executing the command, the system will no longer send log information to any remote syslog server, but only save the log information locally.

【Instance】

```
# configuration sends log information to syslog server 192.168.1.1
on port 8848
Switch> enable
Switch#configure terminal
Switch(config)#log syslog server 192.168.1.1 8848

#Disable the log sending function to the remote syslog server
Switch> enable
Switch#configure terminal
Switch(config)#no log syslog
```

32 NTP Configuration

32.1 NTP server

【Command】

ntp server <A.B.C.D>

no ntp server <A.B.C.D | all>

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: ntp server IP address.

【Description】

ntp server <A.B.C.D>: used to configure the IP address of ntp server and start the ntp service. The default ntp service is not enabled. Only one ntp server is currently supported.

no ntp server <A.B.C.D/all> : delete the configured ntp server IP address and disable the ntp service. A.B.C.D is the address of the NTP server that need to be deleted. Since the system currently supports at most one ntp server IP configuration, the two forms of this command achieve the same effect: delete all ntp server IP addresses and disable the ntp service.

【Instance】

```
# Enable ntp service and configure ntp server ip to 192.168.1.1
Switch> enable
Switch#configure terminal
Switch(config)#ntp server 192.168.1.1
```

```
# Delete all configured ntp server IP and disable the ntp service
(1)
Switch> enable
Switch#configure terminal
Switch(config)#no ntp server 192.168.1.1

# Delete all configured ntp server IP and disable the ntp service
(2)
Switch> enable
Switch#configure terminal
Switch(config)#no ntp server all
```

33 RTC Configuration

33.1 RTC Enable

【Command】

rtc (enable | disable)

【View】

Global configuration mode

【Default Level】

2: Configuration level

【Parameter】

No.

【Description】

rtc enable: used to enable RTC function. RTC(Real-Time Clock) is the pulse generated by the clock circuit composed of crystal oscillator and related circuits on the main board of the device. After the RTC function is enabled, the time information will be obtained from RTC when the device system is started. In addition, RTC clock information will not be lost after the device is powered off.

rtc disable: used to disable RTC function.

【Instance】

Enable RTC function.

```
Switch> enable
Switch#configure terminal
Switch(config)#rtc enable
```

33.2 Display RTC Status

【Command】

show rtc

【View】

Priviledged user mode

【Default Level】

1: View level.

【Parameter】

No.

【Description】

show rtc: view RTC status.

【Instance】

View RTC status.

```
*Switch#show rtc
RTC Status: enable
```

34 Network Diagnose Configuration

34.1 Ping Test

【Command】

ping WORD
ping ip WORD
ping ipv6 WORD
ping

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

WORD: the target IP address that needs to be checked for connectivity.

Ipv6: supported ipv6.

【Description】

None

【Instance】

```
Switch> enable
Switch#ping 192.168.1.188
PING 192.168.1.188 (192.168.1.188): 56 data bytes
64 bytes from 192.168.1.188: seq=0 ttl=128 time=1.493 ms
64 bytes from 192.168.1.188: seq=1 ttl=128 time=13.077 ms

--- 192.168.1.188 ping statistics ---
2 packets transmitted, 2 packets received, 0% packet loss
round-trip min/avg/max = 1.493/7.285/13.077 ms
```

```
*Switch#ping ipv6 fe80::01 (Ipv6 address needs to be configured
to fe80::02/64)
Output Interface: vlanif1
PING fe80::01 (fe80::1): 56 data bytes
64 bytes from fe80::1: seq=0 ttl=128 time=0.536 ms
64 bytes from fe80::1: seq=1 ttl=128 time=0.483 ms

--- fe80::01 ping statistics ---
2 packets transmitted, 2 packets received, 0% packet loss
round-trip min/avg/max = 0.483/0.509/0.536 ms
```

34.2 Traceroute Test

【Command】

traceroute ip WORD

traceroute ipv6 WORD

【View】

Priviledged user mode

【Default Level】

1: View level

【Parameter】

WORD：the connected destination IP address that needs to be checked.

Ipv6: supported ipv6.

【Description】

None

【Instance】

```
Switch> enable
Switch#traceroute ip 192.168.1.254
traceroute to 192.168.1.254 (192.168.1.254), 30 hops max, 38 byte
packets
 1  192.168.1.254 (192.168.1.254)  0.036 ms  0.033 ms  0.013 ms

*Switch#traceroute ipv6 fe80::01 (Ipv6 address needs to be
configured to fe80::02/64 )
Output Interface: vlanif1
traceroute to fe80::01 (fe80::1), 30 hops max, 16 byte packets
 1  fe80::1 (fe80::1)  1.144 ms  0.440 ms  0.346 ms
```

34.3 Port Loopback

【Command】

```
loopback IFNAME internal mac
loopback IFNAME internal phy
no loopback IFNAME internal
```

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

IFNAME: Specify the test port name.

Internal: represents an internal ring test.

mac/phy: the loop need to test is in the MAC layer or phy layer

【Description】

When the port loopback test is enabled, the port link light will be on, no execution will cancel the test, and the port link light will be off.

【Instance】

```
Switch> enable
Switch#loopback ge1 internal mac
Switch#no loopback ge1 internal
```

35 System Maintenance

35.1 Device Information Display

35.1.1 Display System Version

【Command】

show version

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

-

【Description】

show version: the command displays the current system version information.

【Instance】

```
# Display System Current Version Information
Switch> enable
Switch#show version
```

35.1.2 Display Product Information

【Command】

show product-info [<NAME>]

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

NAME: product information name, no parameters by default.

【Description】

show product-info [<NAME>]: the command is used to display the product information value of the given name, and all product information will be displayed when no parameters are provided.

【Instance】

```
# Display Product Information:
Switch> enable
Switch(config)#show product-info
```

35.2 System Software Upgrade

【Command】

copy tftp package <A.B.C.D> <WORD>

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D:tftp server ip address.

WORD: the name of the upgrade file

【Description】

copy tftp package: the command is used to upgrade the system software, in which the parameter A.B.C.D is the IP address of the tftp server, and WORD is the file name of “xxx.bin” for upgrade.

When files are uploaded and downloaded, the tftpd32 software can be used as the tftp server on the PC. When transferring files, make sure the TFTP server is enabled and the correct file path is used.

【Instance】

```
# Upgrade system WWW and product information
Switch> enable
Switch#copy tftp package 192.168.1.1 packetweb.bin

# Upgrade System Software
Switch> enable
Switch#copy tftp package 192.168.1.1 packetapp.bin
```

35.3 Configuration File Import and Export

35.3.1 Import Configuration File

【Command】

copy tftp startup-config <A.B.C.D> <WORD>

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: tftp server ip address

WORD: the name of the upgraded configuration file.

【Description】

copy tftp startup-config: the command is used to upgrade the system configuration file, in which the parameter A.B.C.D is the IP address of the tftp server, and WORD is the name of the configuration file used for the upgrade.

【Instance】

```
# Upgrade System Configuration File
Switch> enable
Switch#copy tftp startup-config 192.168.1.1 SWOS.conf
```

35.3.2 Configure File Export

【Command】

copy flash startup-config <A.B.C.D> (WORD|)

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D：: tftp server ip address.

WORD: the name of the upgraded configuration file.

【Description】

copy tftp startup-config: the command is used to download starp-config files to the tftp server, in which the parameter A.B.C.D is the IP address of the tftp server, and WORD is the file name used when saving to the tftp server.

【Instance】

```
# upload startup-config to tftp server "192.168.1.1" and name it
"SWOS.conf”
Switch> enable
Switch#copy flash startup-config 192.168.1.1 SWOS.conf
```

35.4 Log File Export

【Command】

copy flash logfile <A.B.C.D> (WORD|)

【View】

Priviledged user mode

【Default Level】

2: Configuration level

【Parameter】

A.B.C.D: tftp server ip address.

WORD: the name of the upgraded configuration file.

【Description】

copy flash logfile: the command is used to download logfile to the tftp server, in which the parameter A.B.C.D is the IP address of the tftp server, and WORD is the file name used when saving to the tftp server.

【Instance】

```
# download logfile to tftp server "192.168.1.1" and name it
"message.log”
Switch> enable
Switch#copy flash logfile 192.168.1.1 message.log
```

35.5 Save configuration

【Command】

```
copy running-config startup-config
write
do write
```

【View】

Priviledged user mode

Any

【Default Level】

2: Configuration level

【Parameter】

No.

【Description】

copy running-config startup-config: the command is used to cover the startup-config file with running-config, that is, to save running-config. running-config is the configuration file that is currently running, and startup-config is the configuration file that is currently saved. copy running-config startup-config is to execute a "write" and save the configuration file.

do write: command can perform the save configuration in any mode (except Privileged Exec Mode).

【Instance】

```
# Save running-config
Switch> enable
Switch#copy running-config startup-config
#or
Switch> enable
Switch#configure terminal
Switch(config)#do write
Building configuration...
```

```
[OK]
```

35.6 Reboot the Device

【Command】

reboot

【View】

Priviledged user mode

【Default Level】

1: view level

【Parameter】

None

【Description】

Reboot the Device

【Instance】

```
Switch> enable
Switch#reboot
reboot system? (y/n): y
```

35.7 Restore factory settings

【Command】

erase startup-config
rm startup-config

【View】

Priviledged user mode

【Default Level】

1: Configuration level

【Parameter】

None

【Description】

Delete current configuration file.

【Instance】

```
Switch> enable
```

```
Switch#erase startup-config
erase startup-config ? (y/n): y
Switch#reboot
```