Content

CONTENT 1

CHAPTER 1 BASIC MANAGEMENT CONFIGURATION 1-1

1.1 SWITCH MANAGEMENT 1-1

1.1.1 Management Options 1-1

1.1.2 CLI Interface 1-10

1.2 BASIC SWITCH CONFIGURATION 1-15

1.2.1 Basic Configuration 1-15

1.2.2 Telnet Management 1-16

1.2.3 Configure Switch IP Addresses 1-20

1.2.4 SNMP Configuration 1-21

1.2.5 Switch Upgrade 1-28

1.3 FILE SYSTEM 1-39

1.3.1 Introduction to File Storage Devices 1-39

1.3.2 File System Operation Configuration Task list 1-39

1.3.3 Typical Applications 1-40

1.3.4 Troubleshooting 1-40

1.4 CLUSTER 1-41

1.4.1 Introduction to cluster network management 1-41

1.4.2 Cluster Network Management Configuration Sequence 1-41

1.4.3 Examples of Cluster Administration 1-44

1.4.4 Cluster Administration Troubleshooting 1-45

1.5 USB 1-45

1.5.1 Introduction 1-45

1.5.2 USB Function Configuration List 1-45

1.5.3 USB Function Examples 1-47

1.5.4 USB Function Troubleshooting 1-48

1.6 DEVICE MANAGEMENT 1-48

1.6.1 Device Management Brief 1-48

1.6.2 Device Management Configuration 1-48

CHAPTER 2 LAYER 2 SERVICES CONFIGURATION.....2-1

2.1 PORT CONFIGURATION 2-1

2.1.1 Introduction to Port.....2-1
2.1.2 Network Port Configuration Task List 2-1
2.1.3 Port Configuration Example 2-3
2.1.4 Port Troubleshooting.....2-4

2.2 PORT ISOLATION 2-4

2.2.1 Introduction to Port Isolation Function 2-4
2.2.2 Task Sequence of Port Isolation.....2-4
2.2.3 Port Isolation Function Typical Examples.....2-5

2.3 PORT LOOPBACK DETECTION 2-6

2.3.1 Introduction to Port Loopback Detection Function 2-6
2.3.2 Port Loopback Detection Function Configuration Task List.....2-7
2.3.3 Port Loopback Detection Function Example 2-8
2.3.4 Port Loopback Detection Troubleshooting 2-9

2.4 ULDP 2-10

2.4.1 Introduction to ULDP Function.....2-10
2.4.2 ULDP Configuration Task Sequence.....2-11
2.4.3 ULDP Function Typical Examples 2-13
2.4.4 ULDP Troubleshooting 2-14

2.5 LLDP 2-15

2.5.1 Introduction to LLDP Function 2-15
2.5.2 LLDP Function Configuration Task Sequence.....2-16
2.5.3 LLDP Function Typical Example 2-19
2.5.4 LLDP Function Troubleshooting 2-20

2.6 LLDP-MED 2-20

2.6.1 Introduction to LLDP-MED 2-20
2.6.2 LLDP-MED Configuration Task Sequence.....2-20
2.6.3 LLDP-MED Example.....2-22
2.6.4 LLDP-MED Troubleshooting 2-25

2.7 PORT CHANNEL 2-25

2.7.1 Introduction to Port Channel 2-25
2.7.2 Brief Introduction to LACP.....2-26

2.7.3 Introduction to Load balance....2-28
2.7.4 Port Channel Configuration Task List 2-28
2.7.5 Port Channel Examples....2-31
2.7.6 Troubleshooting....2-34
2.8 MTU....2-34
2.8.1 Introduction to MTU....2-34
2.8.2 MTU Configuration Task Sequence 2-35
2.9 BPDU-TUNNEL....2-35
2.9.1 Introduction to bpdu-tunnel....2-35
2.9.2 bpdu-tunnel Configuration Task List....2-36
2.9.3 Examples of bpdu-tunnel....2-37
2.9.4 bpdu-tunnel Troubleshooting....2-38
2.10 DDM....2-39
2.10.1 Introduction to DDM....2-39
2.10.2 DDM Configuration Task List....2-40
2.10.3 Examples of DDM....2-42
2.10.4 DDM Troubleshooting....2-46
2.11 EFM OAM....2-46
2.11.1 Introduction to EFM OAM....2-46
2.11.2 EFM OAM Configuration....2-49
2.11.3 EFM OAM Example 2-51
2.11.4 EFM OAM Troubleshooting....2-52
2.12 PORT SECURITY....2-52
2.12.1 Introduction to PORT SECURITY 2-52
2.12.2 PORT SECURITY Configuration Task List....2-52
2.12.3 Example of PORT SECURITY 2-53
2.12.4 PORT SECURITY Troubleshooting 2-54
2.13 QSFP+ PORT SPLIT AND COMBINATION 2-54
2.13.1 Introduction to QSFP+ Port Split and Combination Configuration 2-54
2.13.2 QSFP+ Port Configuration 2-54
2.13.3 Typical Case of QSFP+ Port Configuration....2-55
2.13.4 QSFP+ Port Configuration Troubleshooting....2-55
2.14 CFM-OAM CONFIGURATION....2-56

2.14.1 Overview 2-56
2.14.2 CFM OAM Basic Concept 2-57
2.14.3 Introduction of CFM OAM Function 2-59
2.14.4 CFM OAM Basic Function Configuration 2-61
2.14.5 CFM OAM Failure Confirmation 2-66
2.14.6 CFM OAM Failure Orientation 2-67
2.14.7 ULPP Linkage (Selectable) 2-68
2.14.8 Example of Configuration Application 2-71
2.14.9 CFM Troubleshooting 2-74
2.15 VLAN 2-75
2.15.1 Introduction to VLAN 2-75
2.15.2 VLAN Configuration Task List 2-77
2.15.3 Typical VLAN Application 2-79
2.15.4 Typical Application of Hybrid Port 2-80
2.16 GVRP 2-82
2.16.1 Introduction to GVRP 2-82
2.16.2 GVRP Configuration Task List 2-83
2.16.3 Example of GVRP 2-83
2.16.4 GVRP Troubleshooting 2-85
2.17 DOT1Q-TUNNEL 2-85
2.17.1 Introduction to Dot1q-tunnel 2-85
2.17.2 Dot1q-tunnel Configuration 2-86
2.17.3 Typical Applications of the Dot1q-tunnel 2-87
2.17.4 Dot1q-tunnel Troubleshooting 2-88
2.18 VLAN-TRANSLATION 2-88
2.18.1 Introduction to VLAN-translation 2-88
2.18.2 VLAN-translation Configuration 2-88
2.18.3 Typical application of VLAN-translation 2-89
2.18.4 VLAN-translation Troubleshooting 2-90
2.19 DYNAMIC VLAN 2-91
2.19.1 Introduction to Dynamic VLAN 2-91
2.19.2 Dynamic VLAN Configuration 2-91
2.19.3 Typical Application of the Dynamic VLAN 2-93
2.19.4 Dynamic VLAN Troubleshooting 2-94
2.20 VOICE VLAN 2-94

2.20.1 Introduction to Voice VLAN 2-94
2.20.2 Voice VLAN Configuration 2-95
2.20.3 Typical Applications of the Voice VLAN 2-96
2.20.4 Voice VLAN Troubleshooting.....2-97
2.21 SUPER VLAN 2-97
2.21.1 Introduction to Super VLAN.....2-97
2.21.2 Super VLAN Configuration 2-99
2.21.3 Typical Application of Super VLAN..... 2-100
2.21.4 Super VLAN Troubleshooting..... 2-101
2.22 MAC ADDRESS TABLE 2-102
2.22.1 Introduction to MAC Table 2-102
2.22.2 Mac Address Table Configuration Task List 2-104
2.22.3 Typical Configuration Examples 2-105
2.22.4 MAC Table Troubleshooting 2-106
2.23 MAC NOTIFICATION 2-107
2.23.1 Introduction to MAC Notification 2-107
2.23.2 MAC Notification Configuration 2-107
2.23.3 MAC Notification Example 2-108
2.23.4 MAC Notification Troubleshooting..... 2-108
CHAPTER 3 IP SERVICES CONFIGURATION 3-1
3.1 LAYER 3 INTERFACE 3-1
3.1.1 Introduction to Layer 3 Interface.....3-1
3.1.2 Layer 3 Interface Configuration Task List 3-1
3.2 NETWORK MANAGEMENT PORT 3-3
3.2.1 Introduce to IP Network Management Port 3-3
3.2.2 Network Management Port Configuration.....3-3
3.3 IP CONFIGURATION 3-3
3.3.1 Introduction to IPv4, IPv6.....3-3
3.3.2 IP Configuration 3-5
3.3.3 IP Configuration Examples 3-13
3.3.4 IPv6 Troubleshooting 3-18
3.4 IP FORWARDING.....3-18

3.4.1 Introduction to IP Forwarding....3-18
3.4.2 IP Route Aggregation Configuration Task....3-18
3.5 URPF....3-19
3.5.1 Introduction to URPF....3-19
3.5.2 URPF Configuration Task Sequence....3-20
3.5.3 URPF Typical Example....3-21
3.5.4 URPF Troubleshooting....3-21
3.6 ARP....3-22
3.6.1 Introduction to ARP....3-22
3.6.2 ARP Configuration Task List....3-22
3.6.3 ARP Troubleshooting....3-23
3.7 STATION MOVEMENT....3-24
3.7.1 Introduction to l3 station movement....3-24
3.7.2 l3 station movement Configuration Task List....3-24
3.8 ARP SCANNING PREVENTION....3-24
3.8.1 Introduction to ARP Scanning Prevention Function....3-24
3.8.2 ARP Scanning Prevention Configuration Task Sequence....3-25
3.8.3 ARP Scanning Prevention Typical Examples....3-27
3.8.4 ARP Scanning Prevention Troubleshooting Help....3-28
3.9 PREVENT ARP SPOOFING....3-28
3.9.1 Overview....3-28
3.9.2 Prevent ARP Spoofing configuration....3-29
3.9.3 Prevent ARP Spoofing Example....3-30
3.10 ARP GUARD....3-31
3.10.1 Introduction to ARP GUARD....3-31
3.10.2 ARP GUARD Configuration Task List....3-32
3.11 ARP LOCAL PROXY....3-32
3.11.1 Introduction to ARP Local Proxy function....3-32
3.11.2 ARP Local Proxy Function Configuration Task List....3-33
3.11.3 Typical Examples of ARP Local Proxy Function....3-34
3.11.4 ARP Local Proxy Function Troubleshooting....3-34
3.12 GRATUITOUS ARP....3-35
3.12.1 Introduction to Gratuitous ARP....3-35

3.12.2 Gratuitous ARP Configuration Task List 3-36
3.12.3 Gratuitous ARP Configuration Example 3-36
3.12.4 Gratuitous ARP Troubleshooting 3-37
3.13 KEEPALIVE GATEWAY 3-37
3.13.1 Introduction to Keepalive Gateway 3-37
3.13.2 Keepalive Gateway Configuration Task List 3-37
3.13.3 Keepalive Gateway Example 3-38
3.13.4 Kepalive Gteway Troubleshooting 3-39
3.14 DHCP 3-39
3.14.1 Introduction to DHCP 3-39
3.14.2 DHCP Server Configuration 3-40
3.14.3 DHCP Relay Configuration 3-43
3.14.4 DHCP Configuration Examples 3-44
3.14.5 DHCP Troubleshooting 3-46
3.15 DHCP OPTION 82 3-46
3.15.1 Introduction to DHCP option 82 3-46
3.15.2 DHCP option 82 Configuration Task List 3-48
3.15.3 DHCP option 82 Application Examples 3-51
3.15.4 DHCP option 82 Troubleshooting 3-53
3.16 DHCP SNOOPING 3-53
3.16.1 Introduction to DHCP Snooping 3-53
3.16.2 DHCP Snooping Configuration Task Sequence 3-54
3.16.3 DHCP Snooping Typical Application 3-58
3.16.4 DHCP Snooping Troubleshooting Help 3-59
3.17 DHCP OPTION 60 AND OPTION 43 3-59
3.17.1 Introduction to DHCP option 60 and option 43 3-59
3.17.2 DHCP option 60 and option 43 Configuration Task List 3-60
3.17.3 DHCPv6 option 60 and option 43 Example 3-60
3.17.4 DHCP option 60 and option 43 Troubleshooting 3-61
CHAPTER 4 ROUTING PROTOCOL RELATED CONFIGURATION 4-1
4.1 ROUTING PROTOCOL OVERVIEW 4-1

4.1.1 Routing Table....4-1
4.1.2 IP Routing Policy....4-2
4.2 STATIC ROUTE....4-9
4.2.1 Introduction to Static Route....4-9
4.2.2 Introduction to Default Route....4-9
4.2.3 Static Route Configuration Task List....4-9
4.2.4 Static Route Configuration Examples....4-10
4.3 RIP....4-11
4.3.1 Introduction to RIP....4-11
4.3.2 RIP Configuration Task List....4-13
4.3.3 RIP Examples....4-18
4.3.4 RIP Troubleshooting....4-20
4.4 OSPF....4-21
4.4.1 Introduction to OSPF....4-21
4.4.2 OSPF Configuration Task List....4-24
4.4.3 OSPF Examples....4-29
4.4.4 OSPF Troubleshooting....4-37
4.5 BGP....4-38
4.5.1 Introduction to BGP....4-38
4.5.2 BGP Configuration Task List....4-41
4.5.3 Configuration Examples of BGP....4-52
4.5.4 BGP Troubleshooting....4-63
4.6 IPV4 BLACK HOLE ROUTING....4-64
4.6.1 Introduction to Black Hole Routing....4-64
4.6.2 IPv4 Black Hole Routing Configuration Task....4-64
4.6.3 Black Hole Routing Configuration Exmaples....4-64
4.6.4 Black Hole Routing Troubleshooting....4-65
4.7 GRE....4-66
4.7.1 Introduction to GRE Tunnel....4-66
4.7.2 GRE Tunnel Basic Configuration....4-66
4.7.3 Example of GRE Tunnel....4-68
4.7.4 Example of GRE Tunnel Quotes Loopback Group....4-71
4.7.5 GRE Tunnel Troubleshooting....4-75
4.8 ECMP....4-75

4.8.1 Introduction to ECMP 4-75
4.8.2 ECMP Configuration Task List....4-76
4.8.3 ECMP Typical Example....4-76
4.8.4 ECMP Troubleshooting....4-79
4.9 BFD 4-79
4.9.1 Introduction to BFD 4-79
4.9.2 BFD Configuration Task List....4-79
4.9.3 Examples of BFD....4-81
4.9.4 BFD Troubleshooting 4-85
4.10 BGP GR....4-85
4.10.1 Introduction to GR 4-85
4.10.2 GR Configuration Task List....4-86
4.10.3 Typical Example of GR 4-88
4.11 OSPF GR 4-89
4.11.1 Introduction to OSPF GR....4-89
4.11.2 OSPF GR Configuration 4-90
4.11.3 OSPF GR Example 4-91
4.11.4 OSPF GR Troubleshooting....4-92
CHAPTER 5 MULTICAST PROTOCOL RELATED CONFIGURATION 5-1
5.1 MULTICAST 5-1
5.1.1 Introduction to Multicast....5-1
5.1.2 Multicast Address 5-1
5.1.3 IP Multicast Packet Transmission 5-3
5.1.4 IP Multicast Application....5-3
5.2 PIM-DM....5-4
5.2.1 Introduction to PIM-DM 5-4
5.2.2 PIM-DM Configuration Task List....5-5
5.2.3 PIM-DM Configuration Examples 5-7
5.2.4 PIM-DM Troubleshooting....5-7
5.3 PIM-SM....5-8
5.3.1 Introduction to PIM-SM....5-8

5.3.2 PIM-SM Configuration Task List 5-9
5.3.3 PIM-SM Configuration Examples 5-12
5.3.4 PIM-SM Troubleshooting 5-14
5.4 MSDP 5-15
5.4.1 Introduction to MSDP 5-15
5.4.2 Brief Introduction to MSDP Configuration Tasks 5-15
5.4.3 Configuration of MSDP Basic Function 5-16
5.4.4 Configuration of MSDP Entities 5-17
5.4.5 Configuration of Delivery of MSDP Packet 5-18
5.4.6 Configuration of Parameters of SA-cache 5-18
5.4.7 MSDP Configuration Examples 5-19
5.4.8 MSDP Troubleshooting 5-24
5.5 ANYCAST RP 5-25
5.5.1 Introduction to ANYCAST RP 5-25
5.5.2 ANYCAST RP Configuration Task 5-25
5.5.3 ANYCAST RP Configuration Examples 5-28
5.5.4 ANYCAST RP Troubleshooting 5-29
5.6 PIM-SSM 5-29
5.6.1 Introduction to PIM-SSM 5-29
5.6.2 PIM-SSM Configuration Task List 5-30
5.6.3 PIM-SSM Configuration Examples 5-30
5.6.4 PIM-SSM Troubleshooting 5-32
5.7 DVMRP 5-32
5.7.1 Introduction to DVMRP 5-32
5.7.2 DVMRP Configuration Task List 5-33
5.7.3 DVMRP Configuration Examples 5-35
5.7.4 DVMRP Troubleshooting 5-36
5.8 DCSCM 5-36
5.8.1 Introduction to DCSCM 5-36
5.8.2 DCSCM Configuration Task List 5-37
5.8.3 DCSCM Configuration Examples 5-40
5.8.4 DCSCM Troubleshooting 5-41
5.9 IGMP 5-41
5.9.1 Introduction to IGMP 5-41

5.9.2 IGMP Configuration Task List....5-43
5.9.3 IGMP Configuration Examples....5-44
5.9.4 IGMP Troubleshooting....5-45
5.10 IGMP SNOOPING....5-46
5.10.1 Introduction to IGMP Snooping....5-46
5.10.2 IGMP Snooping Configuration Task List....5-46
5.10.3 IGMP Snooping Examples....5-49
5.10.4 IGMP Snooping Troubleshooting....5-51
5.11 IGMP PROXY....5-51
5.11.1 Introduction to IGMP Proxy....5-51
5.11.2 IGMP Proxy Configuration Task List....5-52
5.11.3 IGMP Proxy Examples....5-53
5.11.4 IGMP Proxy Troubleshooting....5-55
5.12 MULTICAST VLAN....5-56
5.12.1 Introductions to Multicast VLAN....5-56
5.12.2 Multicast VLAN Configuration Task List....5-56
5.12.3 Multicast VLAN Examples....5-57
CHAPTER 6 SECURITY FUNCTION CONFIGURATION....6-1
6.1 ACL....6-1
6.1.1 Introduction to ACL....6-1
6.1.2 ACL Configuration Task List....6-2
6.1.3 ACL Example....6-16
6.1.4 ACL Troubleshooting....6-19
6.2 SELF-DEFINED ACL....6-21
6.2.1 Introduction to Self-defined ACL....6-21
6.2.2 Self-defined ACL Configuration....6-22
6.2.3 Self-defined ACL Example....6-24
6.2.4 Self-defined ACL Troubleshooting....6-24
6.3 802.1X....6-25
6.3.1 Introduction to 802.1x....6-25
6.3.2 802.1x Configuration Task List....6-36
6.3.3 802.1x Application Example....6-38
6.3.4 802.1x Troubleshooting....6-39

6.4 THE NUMBER LIMITATION FUNCTION OF MAC AND IP IN PORT, VLAN...6-40

6.4.1 Introduction to the Number Limitation Function of MAC and IP in Port, VLAN.....6-40

6.4.2 The Number Limitation Function of MAC and IP in Port, VLAN Configuration Task Sequence 6-41

6.4.3 The Number Limitation Function of MAC and IP in Port, VLAN Typical Examples.....6-43

6.4.4 The Number Limitation Function of MAC and IP in Port, VLAN Troubleshooting Help 6-44

6.5 AM 6-45

6.5.1 Introduction to AM Function.....6-45

6.5.2 AM Function Configuration Task List 6-45

6.5.3 AM Function Example.....6-46

6.5.4 AM Function Troubleshooting 6-47

6.6 SECURITY FEATURE.....6-47

6.6.1 Introduction to Security Feature 6-47

6.6.2 Security Feature Configuration 6-47

6.6.3 Security Feature Example.....6-50

6.7 TACACS+ 6-50

6.7.1 Introduction to TACACS+.....6-50

6.7.2 TACACS+ Configuration Task List.....6-51

6.7.3 TACACS+ Scenarios Typical Examples.....6-51

6.7.4 TACACS+ Troubleshooting.....6-52

6.8 RADIUS 6-53

6.8.1 Introduction to RADIUS.....6-53

6.8.2 RADIUS Configuration Task List 6-55

6.8.3 RADIUS Typical Examples 6-56

6.8.4 RADIUS Troubleshooting.....6-58

6.9 SSL.....6-58

6.9.1 Introduction to SSL.....6-58

6.9.2 SSL Configuration Task List.....6-60

6.9.3 SSL Typical Example 6-61

6.9.4 SSL Troubleshooting.....6-61

6.10 VLAN-ACL 6-62

6.10.1 Introduction to VLAN-ACL 6-62
6.10.2 VLAN-ACL Configuration Task List 6-62
6.10.3 VLAN-ACL Configuration Example 6-63
6.10.4 VLAN-ACL Troubleshooting 6-65
6.11 CAPTIVE PORTAL AUTHENTICATION 6-65
6.11.1 Captive Portal Authentication Configuration 6-65
6.11.2 Accounting Function Configuration 6-70
6.11.3 Free-resource Configuration 6-73
6.11.4 Authentication White-list Configuration 6-74
6.11.5 Automatic Page Pushing after Successful Authentication (it is not supported currently) 6-75
6.11.6 http-redirect-filter 6-78
6.11.7 Portal Non-perception 6-80
6.11.8 Portal Escaping 6-83
6.12 QOS 6-89
6.12.1 QoS 6-89
6.12.2 PBR 6-102
6.12.3 IPv6 PBR 6-105
6.13 FLOW-BASED REDIRECTION 6-108
6.13.1 Introduction to Flow-based Redirection 6-108
6.13.2 Flow-based Redirection Configuration Task Sequence 6-108
6.13.3 Flow-based Redirection Examples 6-109
6.13.4 Flow-based Redirection Troubleshooting Help 6-109
6.14 EGRESS QOS 6-109
6.14.1 Introduction to Egress QoS 6-109
6.14.2 Egress QoS Configuration 6-111
6.14.3 Egress QoS Examples 6-114
6.14.4 Egress QoS Troubleshooting Help 6-115
6.15 FLEXIBLE QINQ 6-115
6.15.1 Introduction to Flexible QinQ 6-115
6.15.2 Flexible QinQ Configuration Task List 6-116
6.15.3 Flexible QinQ Example 6-117
6.15.4 Flexible QinQ Troubleshooting 6-119
6.16 MPLS QOS 6-120

6.16.1 MPLS QoS Introduction 6-120
6.16.2 MPLS QoS Configuration 6-121
6.16.3 MPLS QoS Examples 6-122
6.16.4 MPLS QoS Troubleshooting Help 6-123
6.17 EGRESS QUEUE SCHEDULING 6-123
6.17.1 Introduction to Egress Queue Scheduling 6-123
6.17.2 Egress Queue Scheduling Configuration 6-127
6.17.3 Egress Queue Scheduling Examples 6-130
6.17.4 Egress Queue Scheduling Troubleshooting 6-131
6.18 VLAN-SHAPING 6-132
6.18.1 Introduction to Vlan-shaping 6-132
6.18.2 Vlan-shaping Configuration Task List 6-135
6.18.3 Vlan-shaping Examples 6-138
6.18.4 Vlan-shaping Troubleshooting 6-143
CHAPTER 7 RELIABILITY CONFIGURATION 7-1
7.1 MSTP 7-1
7.1.1 Introduction to MSTP 7-1
7.1.2 MSTP Configuration Task List 7-3
7.1.3 MSTP Example 7-6
7.1.4 MSTP Troubleshooting 7-11
7.2 ERPS 7-11
7.2.1 Introduction to ERPS 7-11
7.2.2 ERPS Configuration 7-17
7.2.3 ERPS Examples 7-19
7.2.4 ERPS Troubleshooting 7-25
7.3 VRRP 7-25
7.3.1 Introduction to VRRP 7-25
7.3.2 VRRP Configuration Task List 7-26
7.3.3 VRRP Typical Examples 7-28
7.3.4 VRRP Troubleshooting 7-28
7.4 MRPP 7-29
7.4.1 Introduction to MRPP 7-29
7.4.2 MRPP Configuration Task List 7-32

7.4.3 MRPP Typical Scenario 7-33
7.4.4 MRPP Troubleshooting.....7-35
7.5 ULPP 7-36
7.5.1 Introduction to ULPP 7-36
7.5.2 ULPP Configuration Task List 7-38
7.5.3 ULPP Typical Examples.....7-40
7.5.4 ULPP Troubleshooting 7-43
7.6 ULSM.....7-43
7.6.1 Introduction to ULSM.....7-43
7.6.2 ULSM Configuration Task List.....7-44
7.6.3 ULSM Typical Example.....7-45
7.6.4 ULSM Troubleshooting.....7-46
CHAPTER 8 DEBUGGING AND DIAGNOSIS CONFIGURATION 8-1
8.1 MONITOR AND DEBUG 8-1
8.1.1 Ping.....8-1
8.1.2 Ping6.....8-1
8.1.3 Icmp-reply fast.....8-1
8.1.4 Traceroute.....8-1
8.1.5 Traceroute6.....8-2
8.1.6 Show.....8-2
8.1.7 Debug 8-3
8.2 LOGGING 8-3
8.2.1 System Log Introduction.....8-4
8.2.2 System Log Configuration.....8-4
8.2.3 System Log Configuration Example.....8-4
8.3 RELOAD SWITCH AFTER SPECIFIED TIME.....8-4
8.3.1 Introduce to Reload Switch after Specifid Time 8-4
8.3.2 Reload Switch after Specifid Time Task List.....8-4
8.4 DEBUGGING AND DIAGNOSIS FOR PACKETS RECEIVED AND SENT BY CPU 8-5
8.4.1 Introduction to Debugging and Diagnosis for Packets Received and

Sent by CPU 8-5
8.4.2 Debugging and Diagnosis for Packets Received and Sent by CPU
Task List 8-5
8.5 DCP 8-6
8.5.1 Introduction to DCP 8-6
8.5.2 DCP Configuration 8-6
8.5.3 DCP Configuration Examples 8-7
8.5.4 DCP Configuration Troubleshooting 8-8
8.6 COPP 8-9
8.6.1 Introduction to COPP 8-9
8.6.2 COPP Configuration 8-9
8.6.3 COPP Configuration Examples 8-11
8.6.4 COPP Configuration Troubleshooting 8-12
8.7 INFO-CENTER 8-13
8.7.1 Overview 8-13
8.7.2 The content indicates the specific contents of the system
information.Configure Sequence 8-16
8.7.3 Instructions 8-18
8.7.4 Erratum 8-21
8.8 MIRROR 8-21
8.8.1 Introduction to Mirror 8-22
8.8.2 Mirror Configuration Task List 8-22
8.8.3 Mirror Examples 8-23
8.8.4 Device Mirror Troubleshooting 8-23
8.9 RSPAN 8-24
8.9.1 Introduction to RSPAN 8-24
8.9.2 RSPAN Configuration Task List 8-26
8.9.3 Typical Examples of RSPAN 8-27
8.9.4 RSPAN Troubleshooting 8-30
8.10 ERSPAN 8-30
8.10.1 Introduction to ERSPAN 8-30
8.10.2 ERSPAN Configuration Task List 8-31
8.10.3 Typical Examples of ERSPAN 8-31
8.10.4 ERSPAN Troubleshooting 8-33

8.11 sFLOW 8-33
8.11.1 Introduction to sFlow.....8-33
8.11.2 sFlow Configuration Task List.....8-34
8.11.3 sFlow Examples 8-35
8.11.4 sFlow Troubleshooting.....8-36
CHAPTER 9 NETWORK TIME MANAGEMENT CONFIGURATION 9-1
9.1 NTP.....9-1
9.1.1 Introduction to NTP Function 9-1
9.1.2 NTP Function Configuration Task List 9-1
9.1.3 Typical Examples of NTP Function 9-4
9.1.4 NTP Function Troubleshooting 9-4
9.2 SNTP 9-5
9.2.1 Introduction to SNTP 9-5
9.2.2 Typical Examples of SNTP Configuration 9-6
9.3 DNSv4/v6 9-6
9.3.1 Introduction to DNS 9-6
9.3.2 DNSv4/v6 Configuration Task List 9-7
9.3.3 Typical Examples of DNS 9-9
9.3.4 DNS Troubleshooting 9-10
9.4 SUMMER TIME.....9-11
9.4.1 Introduction to Summer Time.....9-11
9.4.2 Summer Time Configuration Task Sequence 9-11
9.4.3 Examples of Summer Time 9-11
9.4.4 Summer Time Troubleshooting.....9-12
CHAPTER 10 VPN CONFIGURATION.....10-1
10.1 MPLS OVERVIEW 10-1
10.1.1 MPLS Overview 10-1
10.2 LDP.....10-7
10.2.1 LDP Introduction 10-7
10.2.2 LDP Configuration 10-16

10.2.3 LDP Typical Instances 10-21
10.2.4 LDP Troubleshooting 10-23
10.3 MPLS VPN 10-24
10.3.1 BGP/MPLS VPN Introduction 10-24
10.3.2 BGP MPLS VPN Configuration 10-34
10.3.3 BGP MPLS VPN Typical Instances 10-40
10.3.4 MPLS BGP VPN Troubleshooting 10-52
10.4 PUBLIC NETWORK ACCESS OF MPLS VPN 10-53
10.4.1 Public Network Access Introduction 10-53
10.4.2 Public Network Access Configuration 10-56
10.4.3 Public Network Access Typical Instances 10-57
10.4.4 Public Network Access Troubleshooting 10-66
10.5 VPLS 10-67
10.5.1 VPLS Overview 10-67
10.5.2 VPLS Configuration 10-74
10.5.3 Typical Examples of VPLS 10-77
10.5.4 VPLS Troubleshooting 10-89
10.6 MAC-IN-MAC 10-90
10.6.1 MAC-in-MAC Overview 10-90
10.6.2 MAC-in-MAC Configuration 10-95
10.6.3 Typical Example of MAC-in-MAC 10-96
10.6.4 MAC-in-MAC Troubleshooting 10-98
CHAPTER 11 VIRTUALIZATION CONFIGURATION 11-1
11.1 VSF 11-1
11.1.1 Overview 11-1
11.1.2 Relevant VSF Configuration 11-7
11.1.3 Typical VSF Example 11-10
11.1.4 VSF Troubleshooting 11-14
CHAPTER 12 OPENFLOW CONFIGURATION 12-1
12.1 OPENFLOW 12-1
12.1.1 Introduction to Openflow 12-1
12.1.2 Openflow Configuration 12-2

12.1.3 Openflow Examples 12-3
12.1.4 Openflow Troubleshooting 12-4
12.2 POX CONTROLLER 12-5
12.2.1 OPENFLOW Configuration (match) 12-5
12.2.2 OPENFLOW Configuration (action) 12-11
12.2.3 OPENFLOW Configuration (examples) 12-14
CHAPTER 13 VXLAN CONFIGURATION 13-1
13.1 VXLAN 13-1
13.1.1 Introduction to VXLAN 13-1
13.1.2 VXLAN Configuration 13-3
13.1.3 VXLAN Example 13-4
13.1.4 VXLAN Troubleshooting 13-6
CHAPTER 14 IPV6 CONFIGURATION 14-1
14.1 DHCPV6 14-1
14.1.1 Introduction to DHCPv6 14-1
14.1.2 DHCPv6 Server Configuration 14-2
14.1.3 DHCPv6 Relay Delegation Configuration 14-3
14.1.4 DHCPv6 Prefix Delegation Server Configuration 14-4
14.1.5 DHCPv6 Prefix Delegation Client Configuration 14-5
14.1.6 DHCPv6 Configuration Examples 14-6
14.1.7 DHCPv6 Troubleshooting 14-9
14.2 DHCPV6 OPTION37, 38 14-10
14.2.1 Introduction to DHCPv6 option37, 38 14-10
14.2.2 DHCPv6 option37, 38 Configuration Task List 14-11
14.2.3 DHCPv6 option37, 38 Examples 14-15
14.2.4 DHCPv6 option37, 38 Troubleshooting 14-18
14.3 PREVENT ND SPOOFING 14-19
14.3.1 Overview 14-19
14.3.2 Prevent ND Spoofing configuration 14-19
14.3.3 Prevent ND Spoofing Example 14-20
14.4 RIPNG 14-20
14.4.1 Introduction to RIPng 14-20

14.4.2 RIPng Configuration Task List 14-21
14.4.3 RIPng Configuration Examples 14-25
14.4.4 RIPng Troubleshooting 14-28
14.5 OSPFv3 14-28
14.5.1 Introduction to OSPFv3 14-28
14.5.2 OSPFv3 Configuration Task List 14-31
14.5.3 OSPFv3 Examples 14-35
14.5.4 OSPFv3 Troubleshooting 14-38
14.6 BGP4+ 14-38
14.6.1 Introduction to MBGP4+ 14-38
14.6.2 MBGP4+ Configuration Task List 14-38
14.6.3 MBGP4+ Examples 14-40
14.6.4 MBGP4+ Troubleshooting 14-41
14.7 IPv6 Black Hole Routing 14-41
14.7.1 Introduction to Black Hole Routing 14-41
14.7.2 IPv6 Black Hole Routing Configuration Task 14-41
14.7.3 Black Hole Routing Configuration Exmaples 14-42
14.7.4 Black Hole Routing Troubleshooting 14-43
14.8 IPv6 Multicast Protocol 14-43
14.8.1 PIM-DM6 14-43
14.8.2 PIM-SM6 14-48
14.8.3 ANYCAST RP v6 14-54
14.8.4 PIM-SSM6 14-59
14.8.5 IPv6 DCSCM 14-62
14.8.6 MLD 14-66
14.8.7 MLD Snooping 14-69
14.9 IPv6 Security RA 14-75
14.9.1 Introduction to IPv6 Security RA 14-75
14.9.2 IPv6 Security RA Configuration Task Sequence 14-75
14.9.3 IPv6 Security RA Typical Examples 14-76
14.9.4 IPv6 Security RA Troubleshooting Help 14-76
14.10 SAVI Configuration 14-77
14.10.1 Introduction to SAVI 14-77
14.10.2 SAVI Configuration 14-77

14.10.3 SAVI Typical Application 14-80
14.10.4 SAVI Troubleshooting..... 14-82
14.11 IPv6 VRRPv3 14-82
14.11.1 Introduction to VRRPv3..... 14-82
14.11.2 VRRPv3 Configuration..... 14-85
14.11.3 VRRPv3 Typical Examples 14-86
14.11.4 VRRPv3 Troubleshooting 14-88

Chapter 1 Basic Management Configuration

1.1 Switch Management

1.1.1 Management Options

After purchasing the switch, the user needs to configure the switch for network management. Switch provides two management options: in-band management and out-of-band management.

1.1.1.1 Out-Of-Band Management

Out-of-band management is the management through Console interface. Generally, the user will use out-of-band management for the initial switch configuration, or when in-band management is not available. For instance, the user must assign an IP address to the switch via the Console interface to be able to access the switch through Telnet.

The procedures for managing the switch via Console interface are listed below:

Step 1: setting up the environment:

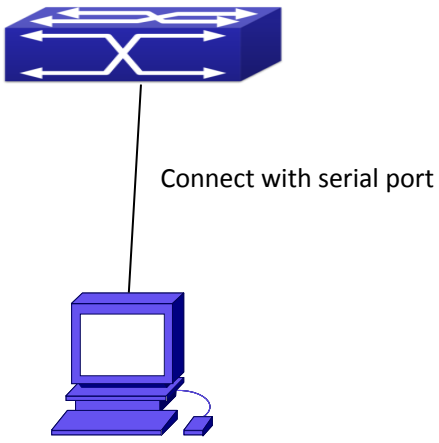


Figure 1-1 Out-of-band Management Configuration Environment

As shown in above, the serial port (RS-232) is connected to the switch with the serial cable provided. The table below lists all the devices used in the connection.

Device Name	Description
PC machine	Has functional keyboard and RS-232, with terminal emulator installed, such as HyperTerminal included in Windows

	9x/NT/2000/XP.
Serial port cable	One end attach to the RS-232 serial port, the other end to the Console port.
Switch	Functional Console port required.

Step 2： Entering the HyperTerminal

Open the HyperTerminal included in Windows after the connection established. The example below is based on the HyperTerminal included in Windows XP.

1) Click Start menu - All Programs -Accessories -Communication - HyperTerminal.

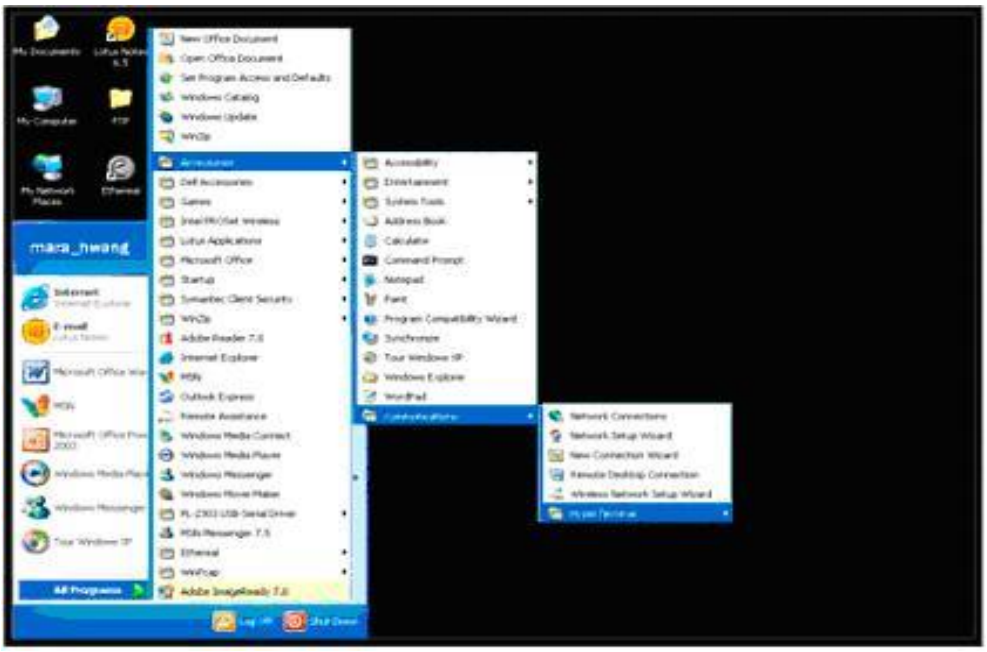


Figure 1-2 Opening Hyper Terminal

2) Type a name for opening HyperTerminal, such as “Switch”.



Figure 1-3 Opening HyperTerminal

3) In the “Connecting using” drop-list, select the RS-232 serial port used by the PC, e.g. COM1, and click “OK”.



Figure 1-4 Opening HyperTerminal

4) COM1 property appears, select “115200” for “Baud rate”, “8” for “Data bits”, “none” for “Parity checksum”, “1” for stop bit and “none” for traffic control; or, you can also click “Restore default” and click “OK”.

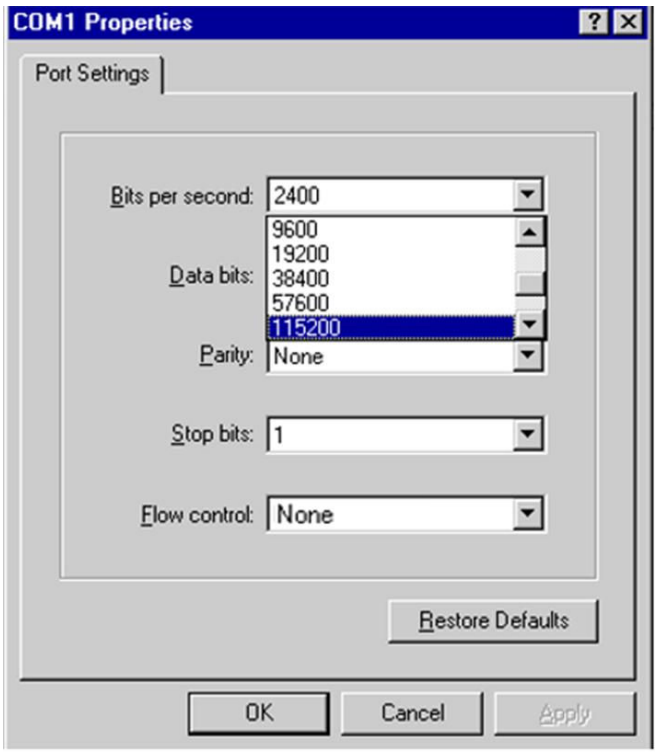


Figure 1-5 Opening HyperTerminal

Step 3: Entering switch CLI interface

Power on the switch, the following appears in the HyperTerminal windows, that is the CLI configuration mode for Switch.

Testing RAM...

0x077C0000 RAM OK

Loading MiniBootROM...

Attaching to file system ...

```
Loading nos.img ...    done.
Booting......
Starting at 0x10000...

Attaching to file system ...
……

--- Performing Power-On Self Tests (POST) ---
DRAM Test.....................PASS!
PCI Device 1 Test.............PASS!
FLASH Test....................PASS!
FAN Test......................PASS!
Done All Pass.
------------------ DONE ----------------------
Current time is SUN JAN 01 00:00:00 2006
……
Switch>
```

The user can now enter commands to manage the switch. For a detailed description for the commands, please refer to the following chapters.

1.1.1.2 In-band Management

In-band management refers to the management by login to the switch using Telnet, or using HTTP, or using SNMP management software to configure the switch. In-band management enables management of the switch for some devices attached to the switch. In the case when in-band management fails due to switch configuration changes, out-of-band management can be used for configuring and managing the switch.

1.1.1.2.1 Management via Telnet

To manage the switch with Telnet, the following conditions should be met:

1) Switch has an IPv4/IPv6 address configured;
2) The host IP address (Telnet client) and the switch's VLAN interface IPv4/IPv6 address is in the same network segment;
3) If 2) is not met, Telnet client can connect to an IPv4/IPv6 address of the switch via other devices, such as a router.

The switch is a Layer 3 switch that can be configured with several IPv4/IPv6 addresses, the configuration method refers to the relative chapter. The following example assumes the shipment status of the switch where only VLAN1 exists in the system.

The following describes the steps for a Telnet client to connect to the switch's VLAN1 interface by Telnet(IPV4 address example):

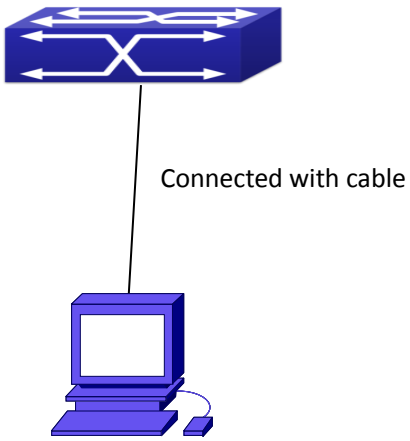


Figure 1-6 Manage the switch by Telnet

Step 1: Configure the IP addresses for the switch and start the Telnet Server function on the switch.

First is the configuration of host IP address. This should be within the same network segment as the switch VLAN1 interface IP address. Suppose the switch VLAN1 interface IP address is 10.1.128.251/24. Then, a possible host IP address is 10.1.128.252/24. Run "ping 10.1.128.251" from the host and verify the result, check for reasons if ping failed.

The IP address configuration commands for VLAN1 interface are listed below. Before in-band management, the switch must be configured with an IP address by out-of-band management (i.e. Console mode), the configuration commands are as follows (All switch configuration prompts are assumed to be "Switch" hereafter if not otherwise specified):

```
Switch>
Switch>enable
Switch#config
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 10.1.128.251 255.255.255.0
Switch(Config-if-Vlan1)#no shutdown
```

To enable the Telnet Server function, users should type the CLI command telnet-server enable in the global mode as below:

```
Switch>enable
Switch#config
Switch(config)# telnet-server enable
```

Step 2: Run Telnet Client program.

Run Telnet client program included in Windows with the specified Telnet target.

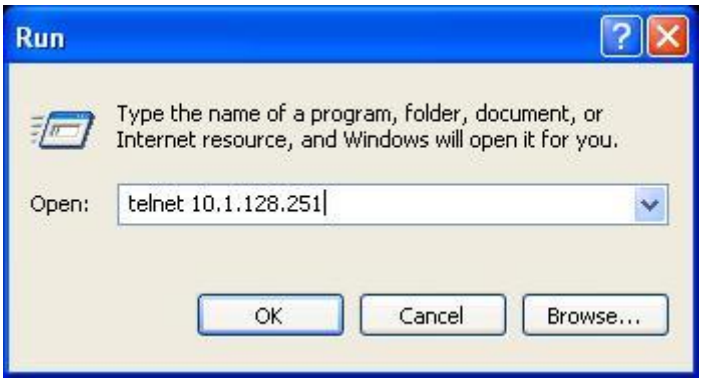


Figure 1-7 Run telnet client program included in Windows

Step 3: Login to the switch.

Login to the Telnet configuration interface. Valid login name and password are required, otherwise the switch will reject Telnet access. This is a method to protect the switch from unauthorized access. As a result, when Telnet is enabled for configuring and managing the switch, username and password for authorized Telnet users must be configured with the following command: username <username> privilege <privilege> [password (0|7) <password>]. To open the local authentication style with the following command: authentication line vty login local. Privilege option must exist and just is 15. Assume an authorized user in the switch has a username of “test”, and password of “test”, the configuration procedure should like the following:

```
Switch>enable
Switch#config
Switch(config)#username test privilege 15 password 0 test
Switch(config)#authentication line vty login local
```

Enter valid login name and password in the Telnet configuration interface, Telnet user will be able to enter the switch’s CLI configuration interface. The commands used in the Telnet CLI interface after login is the same as that in the Console interface.

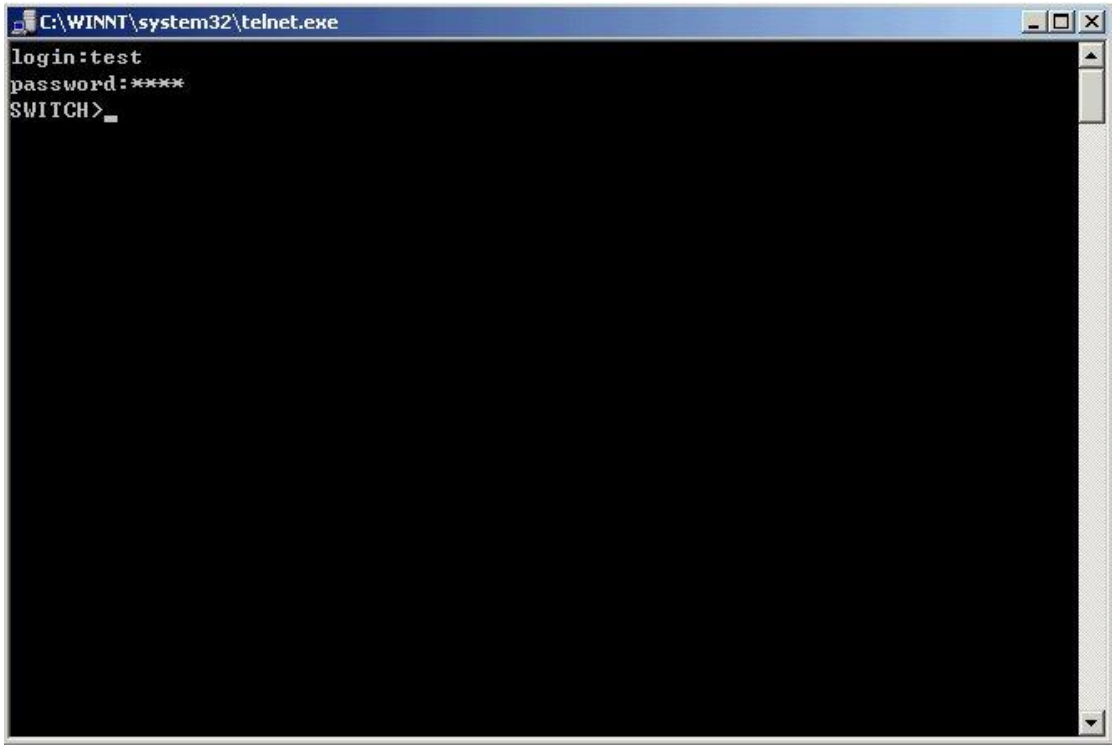


Figure 1-8 Telnet Configuration Interface

1.1.1.2.2 Management via HTTP

To manage the switch via HTTP, the following conditions should be met:

1) Switch has an IPv4/IPv6 address configured;
2) The host IPv4/IPv6 address (HTTP client) and the switch's VLAN interface IPv4/IPv6 address are in the same network segment;
3) If 2) is not met, HTTP client should connect to an IPv4/IPv6 address of the switch via other devices, such as a router.

Similar to management the switch via Telnet, as soon as the host succeeds to ping/ping6 an IPv4/IPv6 address of the switch and to type the right login password, it can access the switch via HTTP. The configuration list is as below:

Step 1: Configure the IP addresses for the switch and start the HTTP server function on the switch.

For configuring the IP address on the switch through out-of-band management, see the telnet management chapter.

To enable the WEB configuration, users should type the CLI command IP http server in the global mode as below:

Switch>enable
Switch#config
Switch(config)#ip http server

Step 2: Run HTTP protocol on the host.

Open the Web browser on the host and type the IP address of the switch, or run directly the HTTP protocol on the Windows. For example, the IP address of the switch is “10.1.128.251”;

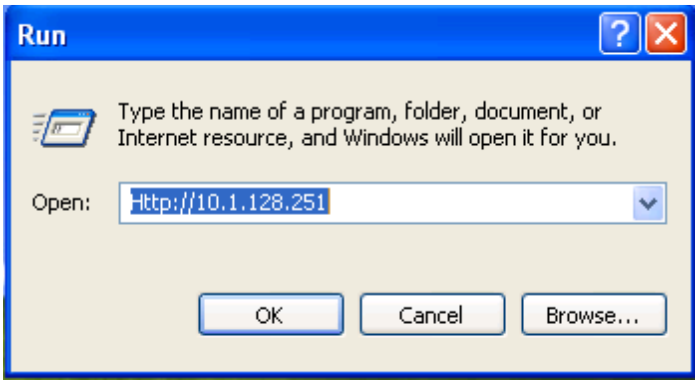


Figure 1-9 Run HTTP Protocol

When accessing a switch with IPv6 address, it is recommended to use the Firefox browser with 1.5 or later version. For example, if the IPv6 address of the switch is 3ffe:506:1:2::3. Input the IPv6 address of the switch is http://[3ffe:506:1:2::3] and the address should draw together with the square brackets.

Step 3: Login to the switch.

Login to the Web configuration interface. Valid login name and password are required, otherwise the switch will reject HTTP access. This is a method to protect the switch from unauthorized access. As a result, when Telnet is enabled for configuring and managing the switch, username and password for authorized Telnet users must be configured with the following command: **username *<username>* privilege *<privilege>* [password (0|7) *<password>*].** To open the local authentication style with the following command: **authentication line web login local.*Privilege*** option must exist and just is 15. Assume an authorized user in the switch has a username of “admin”, and password of “admin”, the configuration procedure should like the following:

Switch>enable
Switch#config
Switch(config)#username admin privilege 15 password 0 admin
Switch(config)#authentication line web login local
The Web login interface of SNR-S4650X-48FQ is as below:

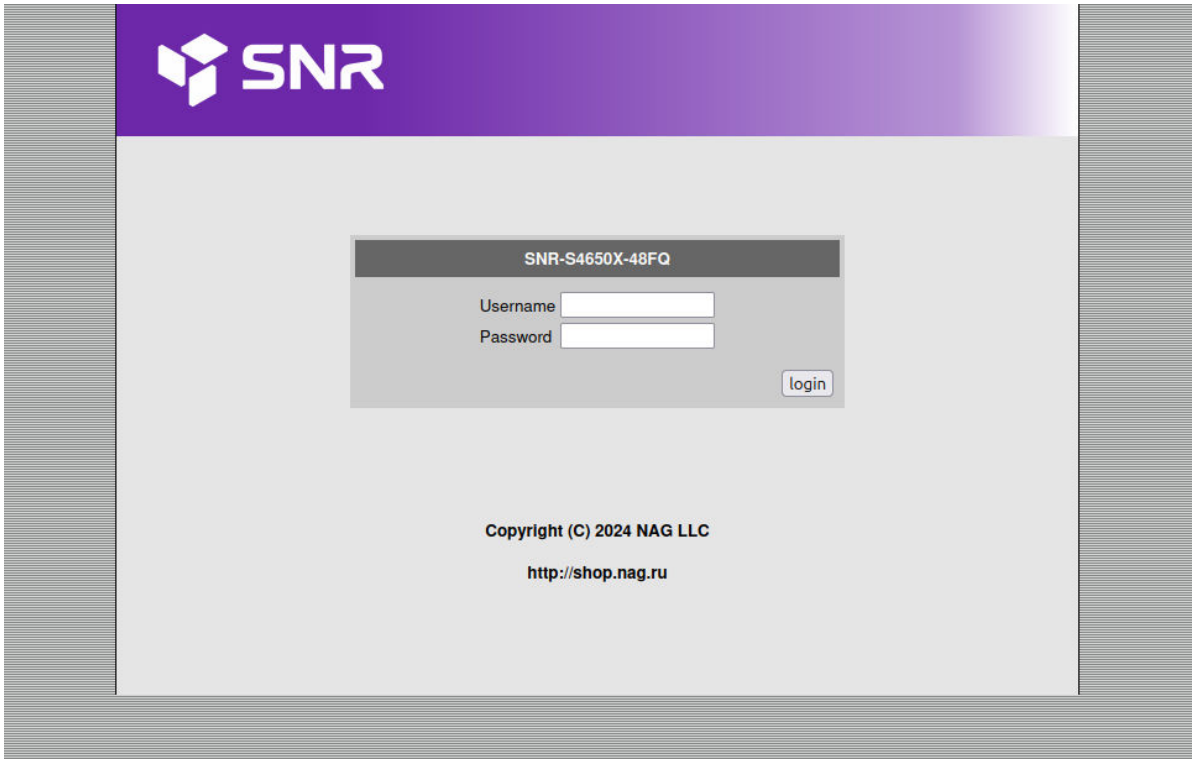


Figure 1-10 Web Login Interface

Input the right username and password, and then the main Web configuration interface is shown as below.

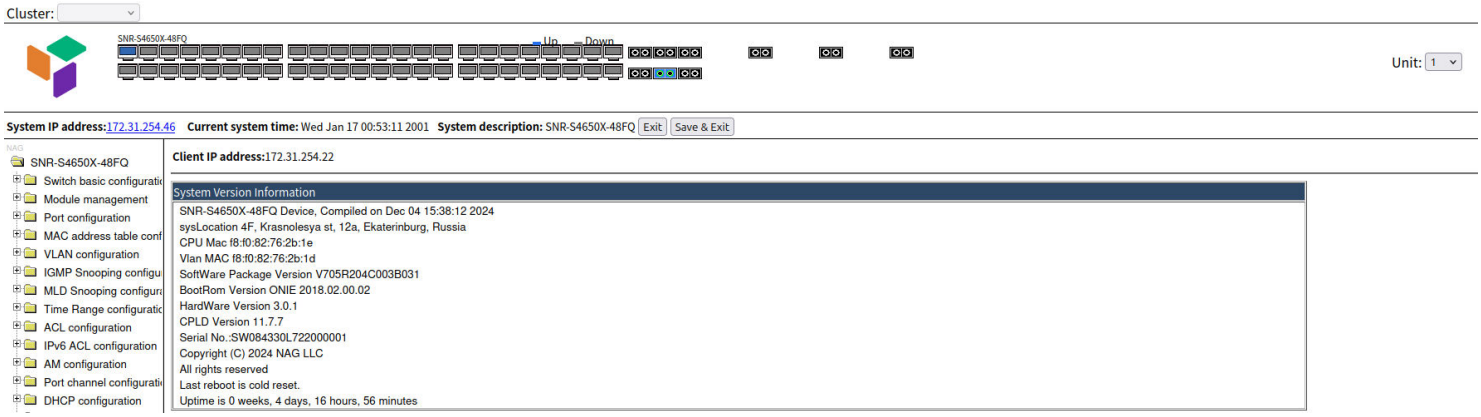


Figure 1-11 Main Web Configuration Interface

Notice: When configure the switch, the name of the switch is composed with English letters.

1.1.1.2.3 Manage the Switch via SNMP Network Management Software

The necessities required by SNMP network management software to manage switches:

1) IP addresses are configured on the switch;
2) The IP address of the client host and that of the VLAN interface on the switch it subordinates to should be in the same segment;
3) If 2) is not met, the client should be able to reach an IP address of the switch through

devices like routers;

4) SNMP should be enabled.

The host with SNMP network management software should be able to ping the IP address of the switch, so that, when running, SNMP network management software will be able to find it and implement read/write operation on it. Details about how to manage switches via SNMP network management software will not be covered in this manual, please refer to “Snmp network management software user manual”.

1.1.2 CLI Interface

The switch provides thress management interface for users: CLI (Command Line Interface) interface, Web interface, Snmp netword management software. We will introduce the CLI interface and Web configuration interface in details, Web interface is familiar with CLI interface function and will not be covered, please refer to “Snmp network management software user manual”.

CLI interface is familiar to most users. As aforementioned, out-of-band management and Telnet login are all performed through CLI interface to manage the switch.

CLI Interface is supported by Shell program, which consists of a set of configuration commands. Those commands are categorized according to their functions in switch configuration and management. Each category represents a different configuration mode. The Shell for the switch is described below:

- Configuration Modes
- Configuration Syntax
- Shortcut keys
- Help function
- Input verification
- Fuzzy match support

1.1.2.1 Configuration Modes

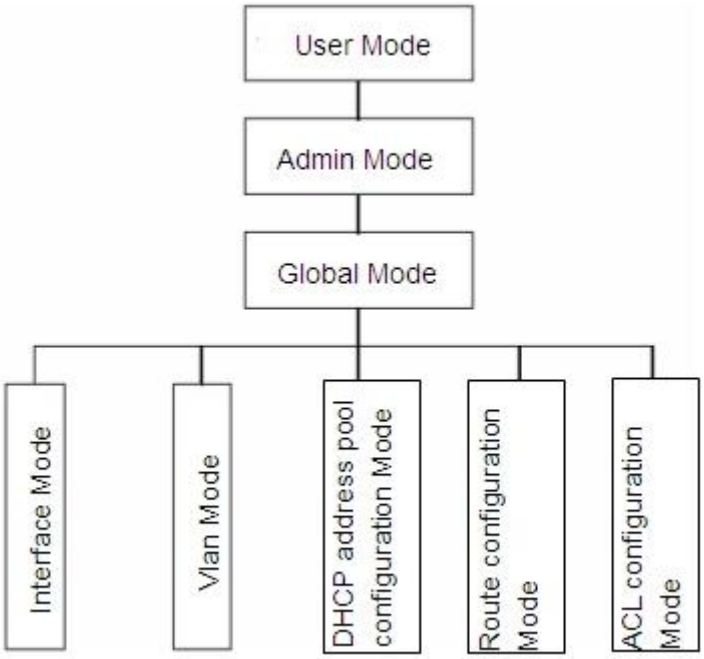


Figure 1-12 Shell Configuration Modes

1.1.2.1.1 User Mode

On entering the CLI interface, entering user entry system first. If as common user, it is defaulted to User Mode. The prompt shown is “Switch>“, the symbol “>“ is the prompt for User Mode. When exit command is run under Admin Mode, it will also return to the User Mode.

Under User Mode, no configuration to the switch is allowed, only clock time and version information of the switch can be queries.

1.1.2.1.2 Admin Mode

To Admin Mode sees the following: In user entry system, if as Admin user, it is defaulted to Admin Mode. Admin Mode prompt “Switch#” can be entered under the User Mode by running the enable command and entering corresponding access levels admin user password, if a password has been set. Or, when exit command is run under Global Mode, it will also return to the Admin Mode. Switch also provides a shortcut key sequence "Ctrl+z”, this allows an easy way to exit to Admin Mode from any configuration mode (except User Mode).

Under Admin Mode, the user can query the switch configuration information, connection status and traffic statistics of all ports; and the user can further enter the Global Mode from Admin Mode to modify all configurations of the switch. For this reason, a password must be set for entering Admin mode to prevent unauthorized access and malicious modification to the switch.

1.1.2.1.3 Global Mode

Type the config command under Admin Mode will enter the Global Mode prompt

"Switch(config)#". Use the exit command under other configuration modes such as Port Mode, VLAN mode will return to Global Mode.

The user can perform global configuration settings under Global Mode, such as MAC Table, Port Mirroring, VLAN creation, IGMP Snooping start and STP, etc. And the user can go further to Port Mode for configuration of all the interfaces.

Interface Mode

Use the interface command under Global Mode can enter the interface mode specified. Switch provides three interface type: 1. VLAN interface; 2. Ethernet port; 3. port-channel, accordingly the three interface configuration modes.

Interface Type	Entry	Operates	Exit
VLAN Interface	Type **interface vlan <*Vlan-id*>** command under Global Mode.	Configure switch IPs, etc	Use the **exit** command to return to Global Mode.
Ethernet Port	Type **interface ethernet <interface-list>** command under Global Mode.	Configure supported duplex mode, speed, etc. of Ethernet Port.	Use the **exit** command to return to Global Mode.
port-channel	Type **interface port-channel <port-channel-number>** command under Global Mode.	Configure port-channel related settings such as duplex mode, speed, etc.	Use the **exit** command to return to Global Mode.

VLAN Mode

Using the vlan <vlan-id> command under Global Mode can enter the corresponding VLAN Mode. Under VLAN Mode the user can configure all member ports of the corresponding VLAN. Run the exit command to exit the VLAN Mode to Global Mode.

DHCP Address Pool Mode

Type the ip dhcp pool <name> command under Global Mode will enter the DHCP Address Pool Mode prompt "Switch(Config-<name>-dhcp)#". DHCP address pool properties can be configured under DHCP Address Pool Mode. Run the exit command to exit the DHCP Address Pool Mode to Global Mode.

Route Mode

Routing Protocol	Entry	Operates	Exit
RIP Routing Protocol	Type **router rip** command under Global Mode.	Configure RIP protocol parameters.	Use the **exit** command to return to Global Mode.
OSPF Routing Protocol	Type **router ospf** command under Global Mode.	Configure OSPF protocol parameters.	Use the **exit** command to return to Global Mode.

BGP Routing Protocol	Type **router bgp <AS mumber>** command under Global Mode.	Configure BGP protocol parameters.	Use the **exit** command to return to Global Mode.

ACL Mode

ACL type	Entry	Operates	Exit
Standard IP ACL Mode	Type **ip access-list standard** command under Global Mode.	Configure parameters for Standard IP ACL Mode.	Use the **exit** command to return to Global Mode.
Extended IP ACL Mode	Type **ip access-list extanded** command under Global Mode.	Configure parameters for Extended IP ACL Mode.	Use the **exit** command to return to Global Mode.

1.1.2.2 Configuration Syntax

Switch provides various configuration commands. Although all the commands are different, they all abide by the syntax for Switch configuration commands. The general commands format of Switch is shown below:

cmdtxt <*variable*> {enum1 | … | enumN } [option1 | … | optionN]

Conventions: **cmdtxt** in bold font indicates a command keyword; ***<variable>*** indicates a variable parameter; **{enum1 | … | enumN }** indicates a mandatory parameter that should be selected from the parameter set **enum1~enumN**; and the square bracket ([ ]) in **[option1 | … | optionN]** indicate an optional parameter. There may be combinations of "< >", "{ }" and "[ ]" in the command line, such as **[*<variable>*]**, **{enum1 *<variable>*| enum2}**, **[option1 [option2]], etc.**

Here are examples for some actual configuration commands:

- show version, no parameters required. This is a command with only a keyword and no parameter, just type in the command to run.
- vlan <vlan-id>, parameter values are required after the keyword.
- firewall {enable | disable}, user can enter firewall enable or firewall disable for this command.
- snmp-server community {ro | rw} <*string*>, the followings are possible:
  snmp-server community ro <string>
  snmp-server community rw <string>

1.1.2.3 Shortcut Key Support

Switch provides several shortcut keys to facilitate user configuration, such as up, down, left, right and Blank Space. If the terminal does not recognize Up and Down keys, ctrl +p and ctrl +n can be used instead.

Key(s)	Function

Back Space	Delete a character before the cursor, and the cursor moves back.	
Up “↑”	Show previous command entered. Up to ten recently entered commands can be shown.	
Down “↓”	Show next command entered. When use the Up key to get previously entered commands, you can use the Down key to return to the next command	
Left “←”	The cursor moves one character to the left.	You can use the Left and Right key to modify an entered command.
Right “→”	The cursor moves one character to the right.	
Ctrl +p	The same as Up key “↑”.	
Ctrl +n	The same as Down key “↓”.	
Ctrl +b	The same as Left key “←”.	
Ctrl +f	The same as Right key “→”.	
Ctrl +z	Return to the Admin Mode directly from the other configuration modes (except User Mode).	
Ctrl +c	Break the ongoing command process, such as ping or other command execution.	
Tab	When a string for a command or keyword is entered, the Tab can be used to complete the command or keyword if there is no conflict.	

1.1.2.4 Help Function

There are two ways in Switch for the user to access help information: the “help” command and the “?”.

Access to Help	Usage and function
Help	Under any command line prompt, type in “help” and press Enter will get a brief description of the associated help system.
“?”	1. Under any command line prompt, enter “?” to get a command list of the current mode and related brief description. 2. Enter a “?” after the command keyword with an embedded space. If the position should be a parameter, a description of that parameter type, scope, etc, will be returned; if the position should be a keyword, then a set of keywords with brief description will be returned; if the output is “<cr>“, then the command is complete, press Enter to run the command. 3. A “?” immediately following a string. This will display all the commands that begin with that string.

1.1.2.5 Input Verification

1.1.2.5.1 Returned Information: success

All commands entered through keyboards undergo syntax check by the Shell. Nothing will be returned if the user entered a correct command under corresponding modes and the execution is successful.

1.1.2.5.2 Returned Information: error

Output error message	Explanation
Unrecognized command or illegal parameter!	The entered command does not exist, or there is error in parameter scope, type or format.
Ambiguous command	At least two interpretations is possible basing on the current input.
Invalid command or parameter	The command is recognized, but no valid parameter record is found.
This command is not exist in current mode	The command is recognized, but this command can not be used under current mode.
Please configure precursor command "*" at first!	The command is recognized, but the prerequisite command has not been configured.
syntax error : missing "" before the end of command line!	Quotation marks are not used in pairs.

1.1.2.5.3 Fuzzy Match Support

Switch shell support fuzzy match in searching command and keyword. Shell will recognize commands or keywords correctly if the entered string causes no conflict.
For example:

1) For command 'show interfaces status ethernet1/0/1', typing 'sh in status ethernet1/0/1' will work.
2) However, for command “show running-config”, the system will report a “> Ambiguous command!” error if only “show r” is entered, as Shell is unable to tell whether it is “show run” or “show running-config”. Therefore, Shell will only recognize the command if “sh ru” is entered.

1.2 Basic Switch Configuration

1.2.1 Basic Configuration

Basic switch configuration includes commands for entering and exiting the admin mode, commands for entering and exiting interface mode, for configuring and displaying the switch clock, for displaying the version information of the switch system, etc.

Command	Explanation
Normal User Mode/ Admin Mode	
enable [*<1-15>*] **disable**	The User uses **enable** command to step into admin mode from normal user mode or modify the privilege level of the users. The **disable** command is for exiting admin mode.
Admin Mode	
config [terminal]	Enter global mode from admin mode.
Various Modes	
exit	Exit current mode and enter previous mode, such as using this command in global mode to go back to admin mode, and back to normal user mode from admin mode.
show privilege	Show privilege of the current users.
Except User Mode/ Admin Mode	
end	Quit current mode and return to Admin mode when not at User Mode/ Admin Mode.
Admin Mode	
clock set *<HH:MM:SS>* [YYYY.MM.DD]	Set system date and time.
show version	Display version information of the switch.
set default	Restore to the factory default.
write	Save current configuration parameters to Flash Memory.
reload	Hot reset the switch.
show cpu usage	Show CPU usage rate.
show cpu utilization	Show current CPU utilization rate.
show memory usage	Show memory usage rate.
Global Mode	
banner motd <LINE> **no banner motd**	Configure the information displayed when the login authentication of a telnet or console user is successful.
web-auth privilege <1-15> **no web-auth privilege**	Configure the level of logging in the switch by web.

1.2.2 Telnet Management

1.2.2.1 Telnet

1.2.2.1.1 Introduction to Telnet

Telnet is a simple remote terminal protocol for remote login. Using Telnet, the user can login to a remote host with its IP address of hostname from his own workstation. Telnet can send the user's keystrokes to the remote host and send the remote host output to the user's screen through TCP connection. This is a transparent service, as to the user, the keyboard and monitor seems to be connected to the remote host directly.

Telnet employs the Client-Server mode, the local system is the Telnet client and the remote host is the Telnet server. Switch can be either the Telnet Server or the Telnet client.

When switch is used as the Telnet server, the user can use the Telnet client program included in Windows or the other operation systems to login to switch, as described earlier in the In-band management section. As a Telnet server, switch allows up to 5 telnet client TCP connections.

And as Telnet client, using telnet command under Admin Mode allows the user to login to the other remote hosts. Switch can only establish TCP connection to one remote host. If a connection to another remote host is desired, the current TCP connection must be dropped.

1.2.2.1.2 Telnet Configuration Task List

1. Configure Telnet Server
2. Telnet to a remote host from the switch.

1. Configure Telnet Server

Command	Explanation
Global Mode	
telnet-server enable **no telnet-server enable**	Enable the Telnet server function in the switch: the no command disables the Telnet function.
username ***<user-name>*** **[privilege** ***<privilege>*****]** **[password [0 | 7]** ***<password>*****]** **no username** ***<username>***	Configure user name and password of the telnet. The no form command deletes the telnet user authorization.
aaa authorization config-commands **no aaa authorization config-commands**	Enable command authorization function for the login user with VTY (login with Telnet and SSH). The no command disables this function. Only enabling this command and configuring command authorization manner, it will request to authorize when executing some command.
authentication securityip ***<ip-addr>*** **no authentication securityip** ***<ip-addr>***	Configure the secure IP address to login to the switch through Telnet: the no command deletes the authorized Telnet secure address.
authentication securityipv6 ***<ipv6-addr>*** **no authentication securityipv6** ***<ipv6-addr>***	Configure IPv6 security address to login to the switch through Telnet; the no command deletes the authorized Telnet security address.

authentication ip access-class {<num-std>|<name>} **no authentication ip access-class**	Binding standard IP ACL protocol to login with Telnet/SSH/Web; the no form command will cancel the binding ACL.
authentication ipv6 access-class {<num-std>|<name>} in **no authentication ipv6 access-class**	Binding standard IPv6 ACL protocol to login with Telnet/SSH/Web; the no form command will cancel the binding ACL.
authentication line {console | vty | web} login method1 [method2 …] **no authentication line {console | vty | web} login**	Configure authentication method list with telnet.
authentication enable method1 [method2 …] **no authentication enable**	Configure the enable authentication method list.
authorization line {console | vty | web} exec method1 [method2 …] **no authorization line {console | vty | web} exec**	Configure the authorization method list with telnet.
authorization line vty command <1-15> {local | radius | tacacs} (none|) **no authorization line vty command <1-15>**	Configure command authorization manner and authorization selection priority of login user with VTY (login with Telnet and SSH). The no command recovers to be default manner.
accounting line {console | vty} command <*1-15*> {start-stop | stop-only | none} method1 [method2…] **no accounting line {console | vty} command <*1-15*>**	Configure the accounting method list.
Admin Mode	
terminal monitor **terminal no monitor**	Display debug information for Telnet client login to the switch; the no command disables the debug information.
show users	Show the user information who logs in through telnet or ssh. It includes line number, user name and user IP.
clear line vty <0-31>	Delete the logged user information on the appointed line, force user to get down the line who logs in through telnet or ssh.

2. Telnet to a remote host from the switch

Command	Explanation
Admin Mode	
telnet [vrf <*vrf-name*>] {<*ip-addr*> | <*ipv6-addr*> / host <*hostname*>} [<*port*>]	Login to a remote host with the Telnet client included in the switch.

1.2.2.2 SSH

1.2.2.2.1 Introduction to SSH

SSH (Secure Shell) is a protocol which ensures a secure remote access connection to network devices. It is based on the reliable TCP/IP protocol. By conducting the mechanism such as key distribution, authentication and encryption between SSH server and SSH client, a secure connection is established. The information transferred on this connection is protected from being intercepted and decrypted. The switch meets the requirements of SSH2.0. It supports SSH2.0 client software such as SSH Secure Client and putty. Users can run the above software to manage the switch remotely.

The switch presently supports RSA authentication, 3DES cryptography protocol and SSH user password authentication etc.

1.2.2.2.2 SSH Server Configuration Task List

Command	Explanation
Global Mode	
ssh-server enable **no ssh-server enable**	Enable SSH function on the switch; the no command disables SSH function.
username ***<username>*** **[privilege** ***<privilege>*****] [password [0 | 7]** ***<password>*****]** **no username** ***<username>***	Configure the username and password of SSH client software for logging on the switch; the no command deletes the username.
ssh-server timeout ***<timeout>*** **no ssh-server timeout**	Configure timeout value for SSH authentication; the no command restores the default timeout value for SSH authentication.
ssh-server authentication-retires ***<authentication-retires>*** **no ssh-server authentication-retries**	Configure the number of times for retrying SSH authentication; the no command restores the default number of times for retrying SSH authentication.
ssh-server host-key create rsa modulus ***<moduls>***	Generate the new RSA host key on the SSH server.
Admin Mode	
terminal monitor **terminal no monitor**	Display SSH debug information on the SSH client side; the no command stops displaying SSH debug information on the SSH client side.
show crypto key	Show the secret key of ssh
crypto key clear rsa	Clear the secret key of ssh.

1.2.2.2.3 Example of SSH Server Configuration

Example1:

Requirement: Enable SSH server on the switch, and run SSH2.0 client software such as

Secure shell client or putty on the terminal. Log on the switch by using the username and password from the client.

Configure the IP address, add SSH user and enable SSH service on the switch. SSH2.0 client can log on the switch by using the username and password to configure the switch.

```
Switch(config)#ssh-server enable
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 100.100.100.200 255.255.255.0
Switch(Config-if-Vlan1)#exit
Switch(config)#username test privilege 15 password 0 test
```

In IPv6 networks, the terminal should run SSH client software which support IPv6, such as putty6. Users should not modify the configuration of the switch except allocating an IPv6 address for the local host.

1.2.3 Configure Switch IP Addresses

All Ethernet ports of switch are default to Data Link layer ports and perform layer 2 forwarding. VLAN interface represent a Layer 3 interface function which can be assigned an IP address, which is also the IP address of the switch. All VLAN interface related configuration commands can be configured under VLAN Mode. Switch provides three IP address configuration methods:

- Manual
- BOOTP
- DHCP

Manual configuration of IP address is assign an IP address manually for the switch.

In BOOTP/DHCP mode, the switch operates as a BOOTP/DHCP client, send broadcast packets of BOOTPRequest to the BOOTP/DHCP servers, and the BOOTP/DHCP servers assign the address on receiving the request. In addition, switch can act as a DHCP server, and dynamically assign network parameters such as IP addresses, gateway addresses and DNS server addresses to DHCP clients DHCP Server configuration is detailed in later chapters.

1.2.3.1 Switch IP Addresses Configuration Task List

1. Enable VLAN port mode
2. Manual configuration
3. BOOTP configuration
4. DHCP configuration

1. Enable VLAN port mode

Command	Explanation
Global Mode	
interface vlan *<vlan-id>* **no interface vlan *<vlan-id>***	Create VLAN interface (layer 3 interface); the no command deletes the VLAN interface.

interface ethernet *<interface-name>*	Enter the network management port configuration mode.

2. Manual configuration

Command	Explanation
VLAN Interface Mode	
ip address *<ip_address>* *<mask>* [secondary] **no ip address *<ip_address>* *<mask>* [secondary]**	Configure IP address of VLAN interface; the no command deletes IP address of VLAN interface.
ipv6 address *<ipv6-address / prefix-length>* [eui-64] **no ipv6 address *<ipv6-address / prefix-length>***	Configure IPv6 address, including aggregation global unicast address, local site address and local link address. The no command deletes IPv6 address.

3. BOOTP configuration

Command	Explanation
VLAN Interface Mode	
ip bootp-client enable **no ip bootp-client enable**	Enable the switch to be a BootP client and obtain IP address and gateway address through BootP negotiation; the no command disables the BootP client function.

4. DHCP configuration

Command	Explanation
VLAN Interface Mode	
ip bootp-client enable **no ip bootp-client enable**	Enable the switch to be a DHCP client and obtain IP address and gateway address through DHCP negotiation; the no command disables the DHCP client function.

1.2.4 SNMP Configuration

1.2.4.1 Introduction to SNMP

SNMP (Simple Network Management Protocol) is a standard network management protocol widely used in computer network management. SNMP is an evolving protocol. SNMP v1 [RFC1157] is the first version of SNMP which is adapted by vast numbers of manufacturers for its simplicity and easy implementation; SNMP v2c is an enhanced version of SNMP v1, which supports layered network management; SNMP v3 strengthens the security by adding USM (User-based Security Mode) and VACM (View-based Access Control Model).

SNMP protocol provides a simple way of exchange network management information between two points in the network. SNMP employs a polling mechanism of message query, and transmits messages through UDP (a connectionless transport layer protocol). Therefore it is well

supported by the existing computer networks.

SNMP protocol employs a station-agent mode. There are two parts in this structure: NMS (Network Management Station) and Agent. NMS is the workstation on which SNMP client program is running. It is the core on the SNMP network management. Agent is the server software runs on the devices which need to be managed. NMS manages all the managed objects through Agents. The switch supports Agent function.

The communication between NMS and Agent functions in Client/Server mode by exchanging standard messages. NMS sends request and the Agent responds. There are seven types of SNMP message:

- Get-Request
- Get-Response
- Get-Next-Request
- Get-Bulk-Request
- Set-Request
- Trap
- Inform-Request

NMS sends queries to the Agent with Get-Request, Get-Next-Request, Get-Bulk-Request and Set-Request messages; and the Agent, upon receiving the requests, replies with Get-Response message. On some special situations, like network device ports are on Up/Down status or the network topology changes, Agents can send Trap messages to NMS to inform the abnormal events. Besides, NMS can also be set to alert to some abnormal events by enabling RMON function. When alert events are triggered, Agents will send Trap messages or log the event according to the settings. Inform-Request is mainly used for inter-NMS communication in the layered network management.

USM ensures the transfer security by well-designed encryption and authentication. USM encrypts the messages according to the user typed password. This mechanism ensures that the messages can't be viewed on transmission. And USM authentication ensures that the messages can't be changed on transmission. USM employs DES-CBC cryptography. And HMAC-MD5 and HMAC-SHA are used for authentication.

VACM is used to classify the users' access permission. It puts the users with the same access permission in the same group. Users can't conduct the operation which is not authorized.

1.2.4.2 Introduction to MIB

The network management information accessed by NMS is well defined and organized in a Management Information Base (MIB). MIB is pre-defined information which can be accessed by network management protocols. It is in layered and structured form. The pre-defined management information can be obtained from monitored network devices. ISO ASN.1 defines a tree structure for MID. Each MIB organizes all the available information with this tree structure. And each node on this tree contains an OID (Object Identifier) and a brief description about the node. OID is a set of integers divided by periods. It identifies the node and can be used to locate the node in a MID tree structure, shown in the figure below:

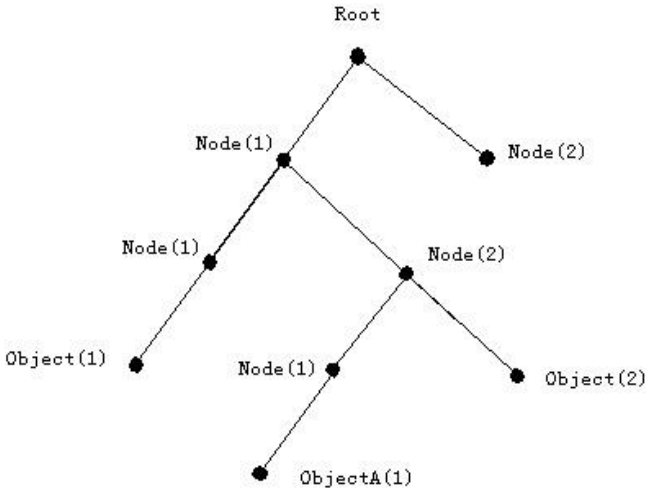


Figure 1-13 ASN.1 Tree Instance

In this figure, the OID of the object A is 1.2.1.1. NMS can locate this object through this unique OID and gets the standard variables of the object. MIB defines a set of standard variables for monitored network devices by following this structure.

If the variable information of Agent MIB needs to be browsed, the MIB browse software needs to be run on the NMS. MIB in the Agent usually consists of public MIB and private MIB. The public MIB contains public network management information that can be accessed by all NMS; private MIB contains specific information which can be viewed and controlled by the support of the manufacturers.

MIB-I [RFC1156] is the first implemented public MIB of SNMP, and is replaced by MIB-II [RFC1213]. MIB-II expands MIB-I and keeps the OID of MIB tree in MIB-I. MIB-II contains sub-trees which are called groups. Objects in those groups cover all the functional domains in network management. NMS obtains the network management information by visiting the MIB of SNMP Agent.

The switch can operate as a SNMP Agent, and supports both SNMP v1/v2c and SNMP v3. The switch supports basic MIB-II, RMON public MIB and other public MID such as BRIDGE MIB. Besides, the switch supports self-defined private MIB.

1.2.4.3 Introduction to RMON

RMON is the most important expansion of the standard SNMP. RMON is a set of MIB definitions, used to define standard network monitor functions and interfaces, enabling the communication between SNMP management terminals and remote monitors. RMON provides a highly efficient method to monitor actions inside the subnets.

MID of RMON consists of 10 groups. The switch supports the most frequently used group 1, 2, 3 and 9:

Statistics: Maintain basic usage and error statistics for each subnet monitored by the Agent.

History: Record periodical statistic samples available from Statistics.

Alarm: Allow management console users to set any count or integer for sample intervals and alert thresholds for RMON Agent records.

Event: A list of all events generated by RMON Agent.

Alarm depends on the implementation of Event. Statistics and History display some current or history subnet statistics. Alarm and Event provide a method to monitor any integer data change in the network, and provide some alerts upon abnormal events (sending Trap or record in logs).

1.2.4.4 SNMP Configuration

1.2.4.4.1 SNMP Configuration Task List

1. Enable or disable SNMP Agent server function
2. Configure SNMP community string
3. Configure IP address of SNMP management base
4. Configure engine ID
5. Configure user
6. Configure group
7. Configure view
8. Configuring TRAP
9. Enable/Disable RMON

1. Enable or disable SNMP Agent server function

Command	Explanation
Global Mode	
snmp-server enabled **no snmp-server enabled**	Enable the SNMP Agent function on the switch; the no command disables the SNMP Agent function on the switch.

2. Configure SNMP community string

Command	Explanation
Global Mode	
snmp-server community {ro | rw} {0 | 7} *<string>* [access {*<num-std>*|*<name>*}] [ipv6-access {*<ipv6-num-std>*|*<ipv6-name>*}] [read *<read-view-name>*] [write *<write-view-name>*] **no snmp-server community *<string>* [access {*<num-std>*|*<name>*}] [ipv6-access {*<ipv6-num-std>*|*<ipv6-name>*}]**	Configure the community string for the switch; the no command deletes the configured community string.

3. Configure IP address of SNMP management station

Command	Explanation
Global Mode	
snmp-server securityip { *<ipv4-address>* /	Configure IPv4/IPv6 security address which is

<ipv6-address> **}** **no snmp-server securityip { *<ipv4-address>* | *<ipv6-address>* }**	allowed to access the switch on the NMS; the no command deletes the configured security address***.***
snmp-server securityip enable **snmp-server securityip disable**	Enable or disable secure IP address check function on the NMS.

4. Configure engine ID

Command	Explanation
Global Mode	
snmp-server engineid *<engine-string>* **no snmp-server engineid**	Configure the local engine ID on the switch. This command is used for SNMP v3.

5. Configure user

Command	Explanation
Global Mode	
snmp-server user *<use-string>* *<group-string>* [{authPriv | authNoPriv} auth {md5 | sha} *<word>*] [access {*<num-std>*|*<name>*}] [ipv6-access {*<ipv6-num-std>*|*<ipv6-name>*}] **no snmp-server user *<user-string>* [access {*<num-std>*|*<name>*}] [ipv6-access {*<ipv6-num-std>*|*<ipv6-name>*}]**	Add a user to a SNMP group. This command is used to configure USM for SNMP v3.

6. Configure group

Command	Explanation
Global Mode	
snmp-server group *<group-string>* {noauthnopriv|authnopriv|authpriv} [[read *<read-string>*] [write *<write-string>*] [notify *<notify-string>*]] [access {*<num-std>*|*<name>*}] [ipv6-access {*<ipv6-num-std>*|*<ipv6-name>*}] **no snmp-server group *<group-string>* {noauthnopriv|authnopriv|authpriv} [access {*<num-std>*|*<name>*}] [ipv6-access {*<ipv6-num-std>*|*<ipv6-name>*}]**	Set the group information on the switch. This command is used to configure VACM for SNMP v3.

7. Configure view

Command	Explanation
Global Mode	
snmp-server view *<view-string>* *<oid-string>* {include|exclude} **no snmp-server view *<view-string>* [*<oid-string>*]**	Configure view on the switch. This command is used for SNMP v3.

8. Configuring TRAP

Command	Explanation
Global Mode	
snmp-server enable traps **no snmp-server enable traps**	Enable the switch to send Trap message. This command is used for SNMP v1/v2/v3.
snmp-server host { <host-ipv4-address> | <host-ipv6-address> } {v1 | v2c | {v3 {noauthnopriv | authnopriv | authpriv}}} *<user-string>* **no snmp-server host { <host-ipv4-address> | <host-ipv6-address> } {v1 | v2c | {v3 {noauthnopriv / authnopriv | authpriv}}} *<user-string>***	Set the host IPv4/IPv6 address which is used to receive SNMP Trap information. For SNMP v1/v2, this command also configures Trap community string; for SNMP v3, this command also configures Trap user name and security level. The "no" form of this command cancels this IPv4 or IPv6 address.
snmp-server trap-source {<ipv4-address> | <ipv6-address>} **no snmp-server trap-source {<ipv4-address> | <ipv6-address>}**	Set the source IPv4 or IPv6 address which is used to send trap packet, the no command deletes the configuration.
Port mode	
[no] switchport updown notification enable	Enable/disable the function of sending the trap message to the port of UP/DOWN event.

9. Enable/Disable RMON

Command	Explanation
Global mode	
rmon enable **no rmon enable**	Enable/disable RMON.

1.2.4.5 Typical SNMP Configuration Examples

The IP address of the NMS is 1.1.1.5; the IP address of the switch (Agent) is 1.1.1.9.

Scenario 1: The NMS network administrative software uses SNMP protocol to obtain data from the switch.

The configuration on the switch is listed below:

Switch(config)#snmp-server enable

Switch(config)#snmp-server community rw private

Switch(config)#snmp-server community ro public

Switch(config)#snmp-server securityip 1.1.1.5

The NMS can use private as the community string to access the switch with read-write permission, or use public as the community string to access the switch with read-only permission.

Scenario 2: NMS will receive Trap messages from the switch (Note: NMS may have community string verification for the Trap messages. In this scenario, the NMS uses a Trap verification

community string of usertrap).
The configuration on the switch is listed below:
Switch(config)#snmp-server enable
Switch(config)#snmp-server host 1.1.1.5 v1 usertrap
Switch(config)#snmp-server enable traps

Scenario 3: NMS uses SNMP v3 to obtain information from the switch.
The configuration on the switch is listed below:
Switch(config)#snmp-server
Switch(config)#snmp-server user tester UserGroup authPriv auth md5 hellotst
Switch(config)#snmp-server group UserGroup AuthPriv read max write max notify max
Switch(config)#snmp-server view max 1 include

Scenario 4: NMS wants to receive the v3Trap messages sent by the switch.
The configuration on the switch is listed below:
Switch(config)#snmp-server enable
Switch(config)#snmp-server host 10.1.1.2 v3 authpriv tester
Switch(config)#snmp-server enable traps
Scenario 5: The IPv6 address of the NMS is 2004:1:2:3::2; the IPv6 address of the switch (Agent) is 2004:1:2:3::1. The NMS network administrative software uses SNMP protocol to obtain data from the switch.
The configuration on the switch is listed below:
Switch(config)#snmp-server enable
Switch(config)#snmp-server community rw private
Switch(config)#snmp-server community ro public
Switch(config)#snmp-server securityip 2004:1:2:3::2

The NMS can use private as the community string to access the switch with read-write permission, or use public as the community string to access the switch with read-only permission.

Scenario 6: NMS will receive Trap messages from the switch (Note: NMS may have community string verification for the Trap messages. In this scenario, the NMS uses a Trap verification community string of usertrap).
The configuration on the switch is listed below:
Switch(config)#snmp-server host 2004:1:2:3::2 v1 usertrap
Switch(config)#snmp-server enable traps

1.2.4.6 SNMP Troubleshooting

When users configure the SNMP, the SNMP server may fail to run properly due to physical connection failure and wrong configuration, etc. Users can troubleshoot the problems by following the guide below:

- ☞ Good condition of the physical connection.
- ☞ Interface and datalink layer protocol is Up (use the “show interface” command), and the

connection between the switch and host can be verified by ping (use “ping” command).

- ☞ The switch enabled SNMP Agent server function (use “snmp-server” command)
- ☞ Secure IP for NMS (use “snmp-server securityip” command) and community string (use “snmp-server community” command) are correctly configured, as any of them fails, SNMP will not be able to communicate with NMS properly.
- ☞ If Trap function is required, remember to enable Trap (use “snmp-server enable traps” command). And remember to properly configure the target host IP address and community string for Trap (use “snmp-server host” command) to ensure Trap message can be sent to the specified host.
- ☞ If RMON function is required, RMON must be enabled first (use “rmon enable” command).
- ☞ Use “show snmp” command to verify sent and received SNMP messages; Use “show snmp status” command to verify SNMP configuration information; Use “debug snmp packet” to enable SNMP debugging function and verify debug information.

If users still can’t solve the SNMP problems, Please contact our technical and service center.

1.2.5 Switch Upgrade

Switch provides two ways for switch upgrade: BootROM upgrade and the TFTP/FTP upgrade under Shell.

1.2.5.1 Switch System Files

The system files include system image file and onie file. The updating of the switch is to update the two files by overwrite the old files with the new ones.

The system image files refers to the compressed files of the switch hardware drivers, and software support program, etc, namely what we usually call the IMG update file. The IMG file can only be saved in the FLASH with a defined name of nos.img

The onie file is the bootloader software maintained by the open source organization. It can be installed in the hardware which conforms to the onie standard.

The update method of the system image file and the onie file is same to each other under the onie mode. The switch supplies the user with two modes of updating system image file: 1. onie mode; 2. TFTP and FTP update at Shell mode. This two update method will be explained in details in following two sections.

1.2.5.2 ONIE Upgrade

There is a method for onie upgrade: TFTP which can be configured at onie command.

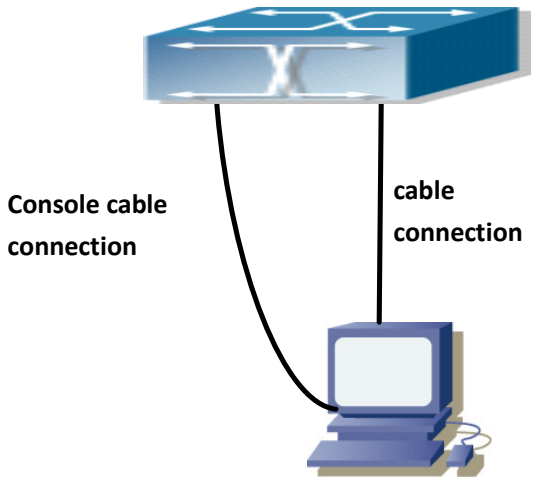


Figure 1-15 Typical topology for switch upgrade in onie mode

The upgrade procedures are listed below:

Step 1:

As shown in the figure, a PC is used as the console for the switch. A console cable is used to connect PC to the management port on the switch. The PC should have TFTP server software installed and has the image file required for the upgrade.

Step 2:

Press the down arrow key when entered into the grub page on switch boot up, choose ONIE mode and press the the down arrow key again, choose Install OS and the operation result is shown below:

GNU GRUB version 2.02~beta2+e4a1fe391

```
 +----------------------------------------------------------------------------------------------------------+
 | nos.img                                                                                                  |
 | nos.img                                                                                                  |
 |*ONIE                                                                                                     |
 |                                                                                                          |
 |                                                                                                          |
 |                                                                                                          |
 |                                                                                                          |
 |                                                                                                          |
 |                                                                                                          |
 |                                                                                                          |
 |                                                                                                          |
 +----------------------------------------------------------------------------------------------------------+
```

GNU GRUB version 2.02~beta2+e4a1fe391

```
  |*ONIE: Install OS                                                                                                  |
  | ONIE: Rescue                                                                                                     |
  | ONIE: Uninstall OS                                                                                               |
  | ONIE: Update ONIE                                                                                                |
  | ONIE: Embed ONIE                                                                                                  |
  |  ACCTON-DIAG                                                                                                      |  |
|                                                                                                                     |
  |                                                                                                                     |
  |                                                                                                                     |
  |                                                                                                                     |
  |                                                                                                                     |
  |
  ||
  |
  +-------------------------------------------------------------------------------------------------------------------+
   ONIE:/
```

Step 3:

Under install os mode, run 'onie-discovery-stop' to disable the discovery action in onie setting up. Then run 'ifconfig eth0' to set the IP address, it should be in the same segment with the address of the connected PC. Suppose the network management port address is 192.168.1.2, and PC address is 192.168.1.66, and the configuration should like:

```
ONIE:/ # onie-discovery-stop
ONIE:/ ifconfig eth0 192.168.1.2
ONIE:/
```

Step 4:

Enable TFTP server in the PC. Run TFTP server program. Before start downloading upgrade file to the switch, verify the connectivity between the server and the switch by ping from the switch. If ping succeeds, run 'onie-nos-install' command in the install os mode from the switch; if it fails, perform troubleshooting to find out the cause. The following is the configuration for the system update image file.

```
ONIE:/ # onie-nos-install tftp://192.168.1.66/nos.onie
discover: installer mode detected.
Stopping: discover... done.
Info: Fetching tftp://192.168.1.66/nos.onie ...
ONIE: Executing installer:
tftp://192.168.1.66/nos.onie
Verifying image checksum ... OK.
```

Preparing image archive ... OK.
SNROS Installer Start...
boardinfo partion exist
SNROS partion exist
mount /dev/sda5 to /boot ok
mount /dev/sda4 to /boardinfo ok
installer will install grub
Success: Support tarball created: /boot/onie-support.tar.bz2
Installing for i386-pc platform.
Installation finished. No error reported.
Installing for i386-pc platform.
grub-install: warning: File system `ext2' doesn't support embedding.
grub-install: warning: Embedding is not possible. GRUB can only be installed in this setup by using blocklists. However, blocklists are UNRELIABLE and their use is discouraged..
Installation finished. No error reported.
/boot/nos.img exists, installer copy /boot/nos.img to /boot/last.img
Rebooting...
ONIE:/ # discover: installer mode detected.
Stopping: discover...start-stop-daemon: warning: killing process 3252: No such process
done.
Stopping: dropbear ssh daemon... done.
Stopping: telnetd... done.
Stopping: syslogd... done.
Info: Unmounting kernel filesystems
The system is going down NOW!
Sent SIGTERM to all processes
Sent SIGKILL toRestarting system.
machine restart

Step 5**:**
Under update onie mode of onie, run 'onie-discovery-stop' to disable the discovery action in onie setting up. Then run 'ifconfig eth0' to set the IP address, it should be in the same segment with the address of the connected PC. At last run 'onie-self-update' to update the onie file.

ONIE:/ # onie-discovery-stop
ONIE:/ # ifconfig eth0 192.168.1.2
ONIE:/ # onie-self-update tftp://192.168.1.66/
ONIE:/ # onie-self-update tftp://192.168.1.66/onie-updater-x86_64-accton

1.2.5.3 FTP/TFTP Upgrade

1.2.5.3.1 Introduction to FTP/TFTP

FTP(File Transfer Protocol)/TFTP(Trivial File Transfer Protocol) are both file transfer protocols that belonging to fourth layer(application layer) of the TCP/IP protocol stack, used for transferring files between hosts, hosts and switches. Both of them transfer files in a client-server model. Their differences are listed below.

FTP builds upon TCP to provide reliable connection-oriented data stream transfer service. However, it does not provide file access authorization and uses simple authentication mechanism (transfers username and password in plain text for authentication). When using FTP to transfer files, two connections need to be established between the client and the server: a management connection and a data connection. A transfer request should be sent by the FTP client to establish management connection on port 21 in the server, and negotiate a data connection through the management connection.

There are two types of data connections: active connection and passive connection.

In active connection, the client transmits its address and port number for data transmission to the server, the management connection maintains until data transfer is complete. Then, using the address and port number provided by the client, the server establishes data connection on port 20 (if not engaged) to transfer data; if port 20 is engaged, the server automatically generates some other port number to establish data connection.

In passive connection, the client, through management connection, notify the server to establish a passive connection. The server then creates its own data listening port and informs the client about the port, and the client establishes data connection to the specified port.

As data connection is established through the specified address and port, there is a third party to provide data connection service.

TFTP builds upon UDP, providing unreliable data stream transfer service with no user authentication or permission-based file access authorization. It ensures correct data transmission by sending and acknowledging mechanism and retransmission of time-out packets. The advantage of TFTP over FTP is that it is a simple and low overhead file transfer service.

Switch can operate as either FTP/TFTP client or server. When switch operates as a FTP/TFTP client, configuration files or system files can be downloaded from the remote FTP/TFTP servers (can be hosts or other switches) without affecting its normal operation. And file list can also be retrieved from the server in ftp client mode. Of course, switch can also upload current configuration files or system files to the remote FTP/TFTP servers (can be hosts or other switches). When switch operates as a FTP/TFTP server, it can provide file upload and download service for authorized FTP/TFTP clients, as file list service as FTP server.

Here are some terms frequently used in FTP/TFTP.

ROM: Short for EPROM, erasable read-only memory. EPROM is repalced by FLASH memory in switch.

SDRAM: RAM memory in the switch, used for system software operation and configuration sequence storage.

FLASH: Flash memory used to save system file and configuration file.

System file: including system image file and boot file.

System image file: refers to the compressed file for switch hardware driver and software support program, usually refer to as IMAGE upgrade file. In switch, the system image file is allowed to save in FLASH only. Switch mandates the name of system image file to be uploaded via FTP in Global Mode to be nos.img, other IMAGE system files will be rejected.

Configuration file: including start up configuration file and running configuration file. The distinction between start up configuration file and running configuration file can facilitate the backup and update of the configurations.

Start up configuration file: refers to the configuration sequence used in switch startup. Startup configuration file stores in nonvolatile storage, corresponding to the so-called configuration save. If the device does not support CF, the configuration file stores in FLASH only, if the device supports CF, the configuration file stores in FLASH or CF, if the device supports multi-config file, names the configuration file to be .cfg file, the default is startup.cfg. If the device does not support multi-config file, mandates the name of startup configuration file to be startup-config.

Running configuration file: refers to the running configuration sequence use in the switch. In switch, the running configuration file stores in the RAM. In the current version, the running configuration sequence running-config can be saved from the RAM to FLASH by **write** command or **copy running-config startup-config** command, so that the running configuration sequence becomes the start up configuration file, which is called configuration save. To prevent illicit file upload and easier configuration, switch mandates the name of running configuration file to be running-config.

Factory configuration file: The configuration file shipped with switch in the name of factory-config. Run **set default** and **write**, and restart the switch, factory configuration file will be loaded to overwrite current start up configuration file.

1.2.5.3.2 FTP/TFTP Configuration

The configurations of switch as FTP and TFTP clients are almost the same, so the configuration procedures for FTP and TFTP are described together in this manual.

1.2.5.3.2.1 FTP/TFTP Configuration Task List

1. FTP/TFTP client configuration
   - （1） Upload/download the configuration file or system file.
   - （2） For FTP client, server file list can be checked.
2. FTP server configuration
   - （1）Start FTP server
   - （2）Configure FTP login username and password
   - （3）Modify FTP server connection idle time
   - （4）Shut down FTP server
3. TFTP server configuration
   - （1）Start TFTP server
   - （2）Configure TFTP server connection idle time
   - （3）Configure retransmission times before timeout for packets without acknowledgement
   - （4）Shut down TFTP server

1. FTP/TFTP client configuration

（1）FTP/TFTP client upload/download file

Command	Explanation
Admin Mode	
copy *<source-url>* *<destination-url>* [ascii | binary]	FTP/TFTP client upload/download file.

（2）For FTP client, server file list can be checked.

Admin Mode	
ftp-dir *<ftpServerUrl>*	For FTP client, server file list can be checked. FtpServerUrl format looks like: ftp: //user: password@IPv4|IPv6 Address.

2. FTP server configuration

（1）Start FTP server

Command	Explanation
Global Mode	
ftp-server enable **no ftp-server enable**	Start FTP server, the no command shuts down FTP server and prevents FTP user from logging in.

（2）Configure FTP login username and password

Command	Explanation
Global Mode	
ip ftp username *<username>* password [0 | 7] *<password>* **no ip ftp username*<username>***	Configure FTP login username and password; this no command will delete the username and password.

（3）Modify FTP server connection idle time

Command	Explanation
Global Mode	
ftp-server timeout *<seconds>*	Set connection idle time.

3. TFTP server configuration

（1）Start TFTP server

Command	Explanation
Global Mode	
tftp-server enable **no tftp-server enable**	Start TFTP server, the no command shuts down TFTP server and prevents TFTP user from logging in.

（2）Modify TFTP server connection idle time

Command	Explanation
Global Mode	
tftp-server retransmission-timeout *<seconds>*	Set maximum retransmission time within timeout interval.

（3）Modify TFTP server connection retransmission time

Command	Explanation

Global Mode	
tftp-server retransmission-number <*number*>	Set the retransmission time for TFTP server.

1.2.5.3.3 FTP/TFTP Configuration Examples

The configuration is same for IPv4 address or IPv6 address. The example only for IPv4 address.

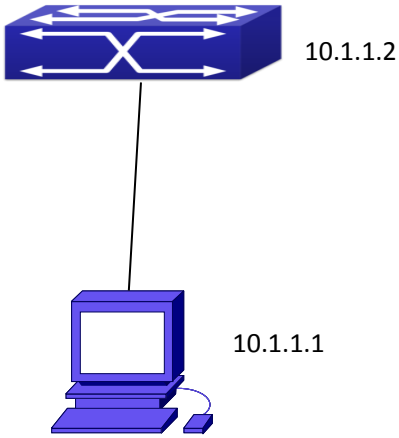


Figure 1-175 Download nos.img file as FTP/TFTP client

Scenario 1: The switch is used as FTP/TFTP client. The switch connects from one of its ports to a computer, which is a FTP/TFTP server with an IP address of 10.1.1.1; the switch acts as a FTP/TFTP client, the IP address of the switch management VLAN is 10.1.1.2. Download “nos.img” file in the computer to the switch.

☞ FTP Configuration

Computer side configuration:

Start the FTP server software on the computer and set the username “Switch”, and the password “superuser”. Place the “12_30_nos.img” file to the appropriate FTP server directory on the computer.

The configuration procedures of the switch are listed below:

```
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-Vlan1)#no shut
Switch(Config-if-Vlan1)#exit
Switch(config)#exit
Switch#copy ftp: //Switch:switch@10.1.1.1/12_30_nos.img nos.img
```

With the above commands, the switch will have the “nos.img” file in the computer downloaded to the FLASH.

☞ TFTP Configuration

Computer side configuration:

Start TFTP server software on the computer and place the “12_30_nos.img” file to the appropriate TFTP server directory on the computer.
The configuration procedures of the switch are listed below:
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-Vlan1)#no shut
Switch(Config-if-Vlan1)#exit
Switch(config)#exit
Switch#copy tftp: //10.1.1.1/12_30_nos.img nos.img

Scenario 2: The switch is used as FTP server. The switch operates as the FTP server and connects from one of its ports to a computer, which is a FTP client. Transfer the “nos.img” file in the switch to the computer and save as 12_25_nos.img.
The configuration procedures of the switch are listed below:
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-Vlan1)#no shut
Switch(Config-if-Vlan1)#exit
Switch(config)#ftp-server enable
Switch(config)# username Admin password 0 superuser

Computer side configuration:
Login to the switch with any FTP client software, with the username “Switch” and password “superuser”, use the command “get nos.img 12_25_nos.img” to download “nos.img” file from the switch to the computer.

Scenario 3: The switch is used as TFTP server. The switch operates as the TFTP server and connects from one of its ports to a computer, which is a TFTP client. Transfer the “nos.img” file in the switch to the computer.
The configuration procedures of the switch are listed below:
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-Vlan1)#no shut
Switch(Config-if-Vlan1)#exit
Switch(config)#tftp-server enable
Computer side configuration:
Login to the switch with any TFTP client software, use the “tftp” command to download “nos.img” file from the switch to the computer.

Scenario 4: Switch acts as FTP client to view file list on the FTP server. Synchronization conditions: The switch connects to a computer by an Ethernet port, the computer is a FTP server with an IP address of 10.1.1.1; the switch acts as a FTP client, and the IP address of the switch management VLAN1 interface is 10.1.1.2.

FTP Configuration:
PC side:
Start the FTP server software on the PC and set the username “Switch”, and the password “superuser”.
Switch:
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-Vlan1)#no shut
Switch(Config-if-Vlan1)#exit
Switch#copy ftp: //Switch: superuser@10.1.1.1
220 Serv-U FTP-Server v2.5 build 6 for WinSock ready...
331 User name okay, need password.
230 User logged in, proceed.
200 PORT Command successful.
150 Opening ASCII mode data connection for /bin/ls.
recv total = 480
nos.img
nos.rom
parsecommandline.cpp
position.doc
qmdict.zip
…(some display omitted here)
show.txt
snmp.TXT
226 Transfer complete.

1.2.5.3.4 FTP/TFTP Troubleshooting

1.2.5.3.4.1 FTP Troubleshooting

When upload/download system file with FTP protocol, the connectivity of the link must be ensured, i.e., use the “Ping” command to verify the connectivity between the FTP client and server before running the FTP program. If ping fails, you will need to check for appropriate troubleshooting information to recover the link connectivity.

☞ The following is what the message displays when files are successfully transferred. Otherwise, please verify link connectivity and retry “copy” command again.
220 Serv-U FTP-Server v2.5 build 6 for WinSock ready...
331 User name okay, need password.
230 User logged in, proceed.
200 PORT Command successful.
nos.img file length = 1526021
read file ok
send file

150 Opening ASCII mode data connection for nos.img.
226 Transfer complete.
close ftp client.

☞ The following is the message displays when files are successfully received. Otherwise, please verify link connectivity and retry “copy” command again.

220 Serv-U FTP-Server v2.5 build 6 for WinSock ready...
331 User name okay, need password.
230 User logged in, proceed.
200 PORT Command successful.
recv total = 1526037
****
write ok
150 Opening ASCII mode data connection for nos.img (1526037 bytes).
226 Transfer complete.

If the switch is upgrading system file or system start up file through FTP, the switch must not be restarted until “close ftp client” or “226 Transfer complete.” is displayed, indicating upgrade is successful, otherwise the switch may be rendered unable to start. If the system file and system start up file upgrade through FTP fails, please try to upgrade again or use the onie mode to upgrade.

1.2.5.3.4.2 TFTP Troubleshooting

When upload/download system file with TFTP protocol, the connectivity of the link must be ensured, i.e., use the “**Ping**” command to verify the connectivity between the TFTP client and server before running the TFTP program. If ping fails, you will need to check for appropriate troubleshooting information to recover the link connectivity.

☞ The following is the message displays when files are successfully transferred. Otherwise, please verify link connectivity and retry “copy” command again.

nos.img file length = 1526021
read file ok
begin to send file, wait...
file transfers complete.
Close tftp client.

☞ The following is the message displays when files are successfully received. Otherwise, please verify link connectivity and retry “copy” command again.

begin to receive file, wait...
recv 1526037
****
write ok
transfer complete
close tftp client.

If the switch is upgrading system file or system start up file through TFTP, the switch must not be restarted until “close tftp client” is displayed, indicating upgrade is successful, otherwise the switch may be rendered unable to start. If the system file and system start up file upgrade

through TFTP fails, please try upgrade again or use the onie mode to upgrade.

1.3 File System

1.3.1 Introduction to File Storage Devices

File storage devices used in switches mainly include FLASH cards. As the most common storage device, FLASH is usually used to store system image files (IMG files), system boot files (ROM files) and system configuration files (CFG files).

Flash can copy, delete, or rename files under Shell or Bootrom mode.

1.3.2 File System Operation Configuration Task list

1.The formatting operation of storage devices
2. Changing the current working directory of the storage device
3. The display operation of the current working directory
4. The display operation of information about a designated file or directory
5. The deletion of a designated file in the file system
6. The renaming operation of files
7. The copying operation of files

1. The formatting operation of storage devices

Command	Explanation
Admin Configuration Mode	
format <*device*>	Format the storage device.

2. Changing the current working directory of the storage device

Command	Explanation
Admin Configuration Mode	
cd <*directory*>	Change the current working directory of the storage device.

3. The display operation of the current working directory

Command	Explanation
Admin Configuration Mode	
pwd	Display the current working directory.

4. The display operation of information about a designated file or directory

Command	Explanation
Admin Configuration Mode	
dir [*WORD*]	Display information about a designated file or directory on the storage device.

5. The deletion of a designated file in the file system

Command	Explanation
Admin Configuration Mode	
delete <*file-url*>	Delete the designated file in the file system.

6. The renaming operation of files

Command	Explanation
Admin Configuration Mode	
rename <*source-file-url*> <*dest-file*>	Change the name of a designated file on the switch to a new one.

7. The copy operation of files

Command	Explanation
Admin Configuration Mode	
copy <*source-file-url* > <*dest-file-url*>	Copy a designated file one the switch and store it as a new one.

1.3.3 Typical Applications

Copy an IMG file flash:/nos.img stored in the FLASH on the boardcard, to cf:/nos-6.1.11.0.img.
The configuration of the switch is as follows:
Switch#copy flash:/nos.img flash:/nos-6.1.11.0.img
Copy flash:/nos.img to flash:/nos-6.1.11.0.img? [Y:N] y
Copyed file flash:/nos.img to flash:/nos-6.1.11.0.img.

1.3.4 Troubleshooting

If errors occur when users try to implement file system operations, please check whether they are caused by the following reasons

- ☞ Whether file names or paths are entered correctly.
- ☞ When renaming a file, whether it is in use or the new file name is already used by an existing file or directory.

1.4 Cluster

1.4.1 Introduction to cluster network management

Cluster network management is an in-band configuration management. Unlike CLI, SNMP and Web Config which implement a direct management of the target switches through a management workstation, cluster network management implements a direct management of the target switches (member switches) through an intermediate switch (commander switch). A commander switch can manage multiple member switches. As soon as a Public IP address is configured in the commander switch, all the member switches which are configured with private IP addresses can be managed remotely. This feature economizes public IP addresses which are short of supply. Cluster network management can dynamically discover cluster feature enabled switches (candidate switches). Network administrators can statically or dynamically add the candidate switches to the cluster which is already established. Accordingly, they can configure and manage the member switches through the commander switch. When the member switches are distributed in various physical locations (such as on the different floors of the same building), cluster network management has obvious advantages. Moreover, cluster network management is an in-band management. The commander switch can communicate with member switches in existing network. There is no need to build a specific network for network management.

Cluster network management has the following features:

- ☞ Save IP addresses
- ☞ Simplify configuration tasks
- ☞ Indifference to network topology and distance limitation
- ☞ Auto detecting and auto establishing
- ☞ With factory default settings, multiple switches can be managed through cluster network management
- ☞ The commander switch can upgrade and configure any member switches in the cluster

1.4.2 Cluster Network Management Configuration Sequence

Cluster Network Management Configuration Sequence:

1. Enable or disable cluster function
2. Create cluster
   1) Configure private IP address pool for member switches of the cluster
   2) Create or delete cluster
   3) Add or remove a member switch
3. Configure attributes of the cluster in the commander switch
   1) Enable or disable automatically adding cluster members
   2) Set automatically added members to manually added ones

3) Set or modify the time interval of keep-alive messages on switches in the cluster.
4) Set or modify the max number of lost keep-alive messages that can be tolerated
5) Clear the list of candidate switches maintained by the switch

4. Configure attributes of the cluster in the candidate switch
   1) Set the time interval of keep-alive messages of the cluster
   2) Set the max number of lost keep-alive messages that can be tolerated in the cluster
5. Remote cluster network management
   1) Remote configuration management
   2) Remotely upgrade member switch
   3) Reboot member switch
6. Manage cluster network with web
   1) Enable http
7. Manage cluster network with snmp
   1) Enable snmp server

1. Enable or disable cluster

Command	Explanation
Global Mode	
cluster run [key <*WORD*>] [vid <*VID*>] **no cluster run**	Enable or disable cluster function in the switch.

2. Create a cluster

Command	Explanation
Global Mode	
cluster ip-pool <*commander-ip*> **no cluster ip-pool**	Configure the private IP address pool for cluster member devices.
cluster commander [<*cluster_name*>] **no cluster commander**	Create or delete a cluster.
cluster member {candidate-sn <*candidate-sn*> | mac-address <*mac-addr*> [id <*member-id*>]} **no cluster member {id <*member-id*> | mac-address <*mac-addr*>}**	Add or remove a member switch.

3. Configure attributes of the cluster in the commander switch

Command	Explanation
Global Mode	
cluster auto-add **no cluster auto-add**	Enable or disable adding newly discovered candidate switch to the cluster.

cluster member auto-to-user	Change automatically added members into manually added ones.
cluster keepalive interval <*second*> **no cluster keepalive interval**	Set the keep-alive interval of the cluster.
cluster keepalive loss-count <*int*> **no cluster keepalive loss-count**	Set the max number of lost keep-alive messages that can be tolerated in the cluster.
Admin mode	
clear cluster nodes [nodes-sn <*candidate-sn-list*> | mac-address <*mac-addr*>]	Clear nodes in the list of candidate switches maintained by the switch.

4. Configure attributes of the cluster in the candidate switch

Command	Explanation
Global Mode	
cluster keepalive interval <*second*> **no cluster keepalive interval**	Set the keep-alive interval of the cluster.
cluster keepalive loss-count <*int*> **no cluster keepalive loss-count**	Set the max number of lost keep-alive messages that can be tolerated in the clusters.

5. Remote cluster network management

Command	Explanation
Admin Mode	
rcommand member <*member-id*>	In the commander switch, this command is used to configure and manage member switches.
rcommand commander	In the member switch, this command is used to configure the commander switch.
cluster reset member [id <*member-id*> | mac-address <*mac-addr*>]	In the commander switch, this command is used to reset the member switch.
cluster update member <*member-id*> <*src-url*> <*dst-filename*>[ascii | binary]	In the commander switch, this command is used to remotely upgrade the member switch. It can only upgrade nos.img file.

6. Manage cluster network with web

Command	Explanation
Global Mode	

ip http server	Enable http function in commander switch and member switch. Notice: must insure the http function be enabled in member switch when commander switch visiting member switch by web. The commander switch visit member switch via beat member node in member cluster topology.

7. Manage cluster network with snmp

Command	Explanation
Global Mode	
snmp-server enable	Enable snmp server function in commander switch and member switch. Notice: must insure the snmp server function be enabled in member switch when commander switch visiting member switch by snmp. The commander switch visit member switch via configure character string <commander-community>@sw<member id>.

1.4.3 Examples of Cluster Administration

Scenario:

The four switches SW1-SW4, amongst the SW1 is the command switch and other switches are member switch. The SW2 and SW4 is directly connected with the command switch, SW3 connects to the command switch through SW2.

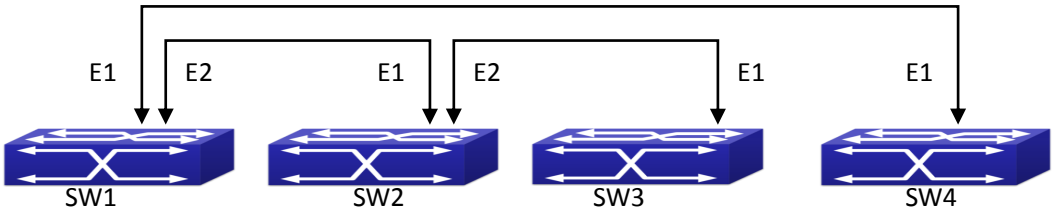


Figure 1-186 Examples of Cluster

Configuration Procedure

1. Configure the command switch

Configuration of SW1:

Switch(config)#cluster run

Switch(config)#cluster ip-pool 10.2.3.4
Switch(config)#cluster commander 5526
Switch(config)#cluster auto-add

2. Configure the member switch
Configuration of SW2-SW4
Switch(config)#cluster run

1.4.4 Cluster Administration Troubleshooting

When encountering problems in applying the cluster admin, please check the following possible causes:

- If the command switch is correctly configured and the auto adding function (cluster auto-add) is enabled. If the ports connected the command switch and member switch belongs to the cluster vlan.
- After cluster commander is enabled in VLAN1 of the command switch, please don't enable a routing protocol (RIP, OSPF, BGP) in this VLAN in order to prevent the routing protocol from broadcasting the private cluster addresses in this VLAN to other switches and cause routing loops.
- Whether the connection between the command switch and the member switch is correct. We can use the debug cluster packets to check if the command and the member switches can receive and process related cluster admin packets correctly.

1.5 USB

1.5.1 Introduction

When there is USB device inserted or pulled out, the switch can detect that information of USB hot inserting and pulling out and the switch will mount or uninstall the USB device.

When there is USB device inserted, the switch will mount the USB file system. It can read, copy, delete, rename the files in USB, and it can also recover the configuration, download the files and save the files.

This device supports the flow-passed warning function of the USB. When the actual electric current exceeds the rated current of the device, the switch will prompt user that the temperature is too high. And then, there is the danger of burning out the device if the USB is inserted.

1.5.2 USB Function Configuration List

1. Mount the USB device and enter in the USB letter

2. Show the USB letter information
3. Copy the source file to be the destination file
4. Delete the file content
5. Rename the file name
6. Update the config file under the USB letter to the switch
7. Update the bootrom file under the USB letter to the switch
8. Update the img file under the USB letter to the switch
9. Create the content
10. Delete the existed content
11. Uninstall the USB device

1. Mount the USB device and enter in the USB letter

Command	Explanation
Admin Mode	
cd usb:	Enter in the USB letter.

2. Show the USB letter information

Command	Explanation
Admin Mode	
dir	Show the USB letter information.

3. Copy the source file to be the destination file

Command	Explanation
Admin Mode	
copy source destination	Copy the source file to be the destination file.

4. Delete the file content

Command	Explanation
Admin Mode	
delete filename	Delete the file.

5. Rename the file name

Command	Explanation
Admin Mode	
rename source destitation	Rename the source file name to be the destination file name.

6. Update the config file under the USB letter to the switch

Command	Explanation

Admin Mode	
copy usb:/startup.cfg startup.cfg	Update the config file under the USB letter to the switch. The reverse transmission is supported: copy startup.cfg usb:/startup.cfg

7. Update the bootrom file under the USB letter to the switch

Command	Explanation
Admin Mode	
copy usb:/boot.rom boot.rom	Update the bootrom file under the USB letter to the switch. The reverse transmission is supported: copy boot.rom usb:/boot.rom

8. Update the img file under the USB letter to the switch

Command	Explanation
Admin Mode	
copy usb:/nos.img nos.img	Update the img file under the USB letter to the switch. The reverse transmission is supported: copy nos.img usb:/nos.img

9. Create the content

Command	Explanation
Admin Mode	
mkdir	Create the content.

10. Delete the existed content

Command	Explanation
Admin Mode	
rmdir	Delete the existed content.

1.5.3 USB Function Examples

Delete source1.txt in the usb letter, and rename tt.txt in the usb letter as tttt.txt.
Create the new content of sw1 in the usb letter.

Switch#delete source1.txt
Switch#rename usb:/tt.txt usb:/tttt.txt
Switch#mkdir sw1

1.5.4 USB Function Troubleshooting

- ☞ Currently, only the USB device mounting and uninstalling under the CLI user operation page mode is supported. This command is not supported under the non-CLI user operation page mode.
- ☞ Make sure the switch is power-on and the USB device is inserted correctly. The file content in the USB device will mount to the file system of the switch automatically.
- ☞ For the reading and writing function of USB, the hot inserting and pulling out are not supported currently.
- ☞ Insert the USB device, the file content will not mount to the switch file system, this function does not support the Chinese recognition and displaying of the file content.
- ☞ Input dir command directly, show the file content under the flash letter as default. If user want it to show the file information of the usb, input **cd usb:** to enter the usb letter, and then input **dir** to show the file information. User can also use the absolute path and input **dir usb:** to show the file content.
- ☞ This command does not support the big file showing currently.

1.6 Device Management

1.6.1 Device Management Brief

The device management function of switch provides information about line card status, line card operation debugging, power supply and fan status. This function enables the maintenance and management of the physical devices and restart of the switch and line cards, and hot swapping of the cards. Switch supports dual-master mode. If 2 master management cards are present in the system, the master control board in the smaller slot number becomes the Active Master and the other board becomes the Standby Master.

1.6.2 Device Management Configuration

1.6.2.1 Monitor and Debug Task

1. Display the chip information
2. Display information of the fan status
3. Display information of the power status

1. Display the chip information

Command	Explanation

Admin Mode	
show [member <member-id>] slot <slot-id>	Show basic information of each chip.

2. Display the information of the fan status

Command	Explanation
Admin Mode	
show fan	Shows whether the fan tray is in place and its running status, and shows the speed of the fan.

3. Display the information of the power status

Command	Explanation
Admin Mode	
show power	Shows if the power supply is in place and its running status.

Chapter 2 Layer 2 services Configuration

2.1 Port Configuration

2.1.1 Introduction to Port

Switch contains 48 10G optical ports and 4 4x10G optical ports, each 4x10G optical port can be distributed into 4 10G optical ports and they support the same configuration and property with the ordinary SFP+ physical ports. After the 4x10G port distributed, a special one-to-four cable can be used to connect them

If the user needs to configure some network ports, he/she can use the interface ethernet <interface-list> command to enter the appropriate Ethernet port configuration mode, where <interface-list> stands for one or more ports. If <interface-list> contains multiple ports, special characters such as ';' or '-' can be used to separate ports, ';' is used for discrete port numbers and '-' is used for consecutive port numbers. Suppose an operation should be performed on ports 2,3,4,5 the command would look like: interface ethernet 1/0/2-5. Port speed, duplex mode and traffic control can be configured under Ethernet Port Mode causing the performance of the corresponding network ports to change accordingly.

2.1.2 Network Port Configuration Task List

1. Enter the network port configuration mode
2. Configure the properties for the network ports
   (1) Enable/Disable ports
   (2) Configure port names
   (3) Configure port speed and duplex mode
   (4) Configure bandwidth control
   (5) Configure traffic control
   (6) Enable/Disable port loopback function
   (7) Configure broadcast storm control function for the switch
   (8) Configure scan port mode
   (9) Configure rate-violation control of the port
   (10) Configure interval of port-rate-statistics

1. Enter the Ethernet port configuration mode

Command	Explanation

Global Mode	
interface ethernet *<interface-list>*	Enters the network port configuration mode.

. Configure the properties for the Ethernet ports

Command	Explanation
Port Mode	
shutdown **no shutdown**	Enables/Disables specified ports.
description <string> **no description**	Specifies or cancels the name of specified ports.
speed-duplex {auto [10 [100 [1000]] [auto | full | half |]] | force10-half | force10-full | force100-half | force100-full | force100-fx [module-type {auto-detected | no-phy-integrated | phy-integrated}] | {{force1g-half | force1g-full} [nonegotiate [master | slave]]}| force10g-full} **no speed-duplex**	Sets port speed and duplex mode of 100/1000Base-TX or 100Base-FX ports. The no format of this command restores the default setting, i.e., negotiates speed and duplex mode automatically.
bandwidth control <bandwidth> [both | receive | transmit] **no bandwidth control**	Sets or cancels the bandwidth used for incoming/outgoing traffic for specified ports.
flow control **no flow** control	Enables/Disables traffic control function for specified ports.
loopback **no *loopback***	Enables/Disables loopback test function for specified ports.
storm-control {*unicast* | broadcast | multicast} <packets>	Enables the storm control function for broadcasts, multicasts and unicasts with unknown destinations (short for broadcast), and sets the allowed broadcast packet number; the no format of this command disables the broadcast storm control function.
port-scan-mode {interrupt | poll} **no port-scan-mode**	Configure port-scan-mode as interrupt or poll mode, the no command restores the default port-scan-mode.
rate-violation <200-2000000> [recovery <0-86400>] ***no* rate-violation**	Set the max packet reception rate of a port. If the rate of the received packet violates the packet reception rate, shut down this port and configure the recovery time, the default is 300s. The no command will disable the rate-violation function of a port.
Interface mode <auto-detect | cr | cr4 | sfi | sr | sr4 | xfi >	Configure the mode of 40G port and the default mode is auto-detect.
Global Mode	

hardware profile mode 4x10G **no hardware profile mode 4x10G**	Configure one 40G port as four 10G ports. The no command disables the 4x10G mode.
port-rate-statistics interval **<interval -value>**	Configure the interval of port-rate-statistics.

2.1.3 Port Configuration Example

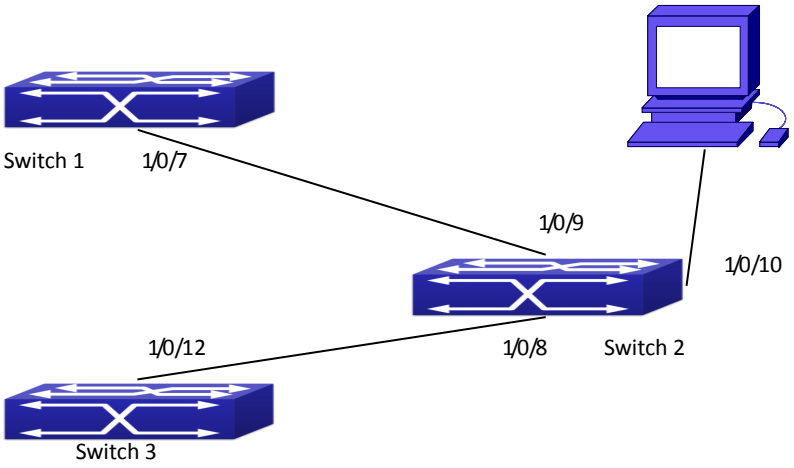


Figure 2-1 Port Configuration Example

No VLAN has been configured in the switches, default VLAN1 is used.

Switch	Port	Property
Switch1	1/0/7	Ingress bandwidth limit: 50 M
Switch2	1/0/8	Mirror source port
	1/0/9	1000Mbps full, mirror source port
	1/0/10	10000Mbps full, mirror destination port
Switch3	1/0/12	1000Mbps full

The configurations are listed below:

Switch1:

Switch1(config)#interface ethernet 1/0/7

Switch1(Config-If-Ethernet1/0/7)#bandwidth control 50000 both

Switch2**:**

Switch2(config)#interface ethernet 1/0/9

Switch2(Config-If-Ethernet1/0/9)#speed-duplex forceg-full

Switch2(Config-If-Ethernet1/0/9)#exit

Switch2(config)#interface ethernet 1/0/10

Switch2(Config-If-Ethernet1/0/10)#speed-duplex force10g-full

Switch2(Config-If-Ethernet1/0/10)#exit

Switch2(config)#monitor session 1 source interface ethernet 1/0/8;1/0/9

Switch2(config)#monitor session 1 destination interface ethernet 1/0/10

Switch3:
Switch3(config)#interface ethernet 1/0/12
Switch3(Config-If-Ethernet1/0/12)#speed-duplex force1g-full
Switch3(Config-If-Ethernet1/0/12)#exit

2.1.4 Port Troubleshooting

Here are some situations that frequently occurs in port configuration and the advised solutions:

- Two connected fiber interfaces won't link up if one interface is set to auto-negotiation but the other to forced speed/duplex. This is determined by IEEE 802.3.
- The following combinations are not recommended: enabling traffic control as well as setting multicast limiting for the same port; setting broadcast, multicast and unknown destination unicast control as well as port bandwidth limiting for the same port. If such combinations are set, the port throughput may fall below the expected performance.
-

2.2 Port Isolation

2.2.1 Introduction to Port Isolation Function

Port isolation is an independent port-based function working in an inter-port way, which isolates flows of different ports from each other. With the help of port isolation, users can isolate ports within a VLAN to save VLAN resources and enhance network security. After this function is configured, the ports in a port isolation group will be isolated from each other, while ports belonging to different isolation groups or no such group can forward data to one another normally. No more than 16 port isolation groups can a switch have.

2.2.2 Task Sequence of Port Isolation

1. Create an isolate port group
2. Add Ethernet ports into the group
3. Specify the flow to be isolated
4. Display the configuration of port isolation

1. Create an isolate port group

Command	Explanation
Global Mode	

isolate-port group <*WORD*> **no isolate-port group <*WORD*>**	Set a port isolation group; the no operation of this command will delete the port isolation group.

2. Add Ethernet ports into the group

Command	Explanation
Global Mode	
isolate-port group <*WORD*> switchport interface [ethernet | port-channel] <*IFNAME*> **no isolate-port group <*WORD*> switchport interface [ethernet | port-channel] <*IFNAME*>**	Add one port or a group of ports into a port isolation group to isolate, which will become isolated from the other ports in the group; the no operation of this command will remove one port or a group of ports out of a port isolation group.

3. Specify the flow to be isolated

Command	Explanation
Global Mode	
isolate-port apply [<l2|*l3*|*all*>]	Apply the port isolation configuration to isolate layer-2 flows, layer-3 flows or all flows.

4. Display the configuration of port isolation

Command	Explanation
Admin Mode and global Mode	
show isolate-port group [<*WORD*>]	Display the configuration of port isolation, including all configured port isolation groups and Ethernet ports in each group.

2.2.3 Port Isolation Function Typical Examples

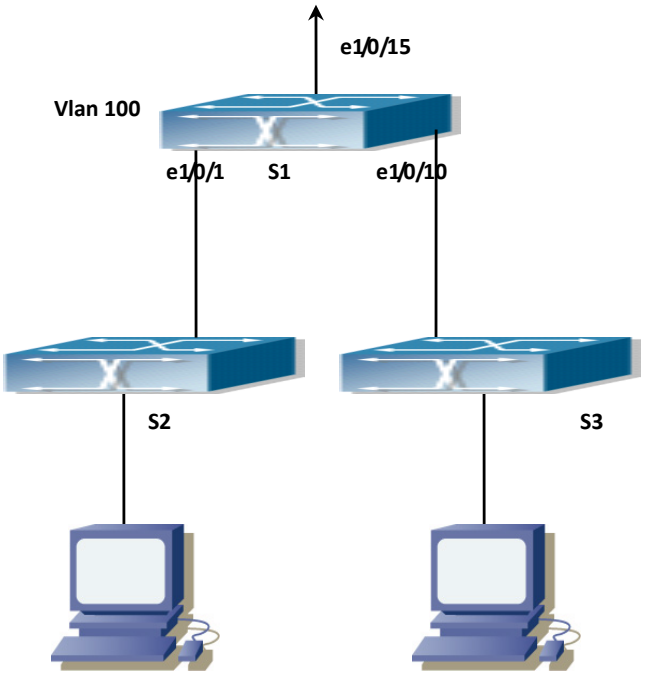


Figure 2-2 Typical example of port isolation function

The topology and configuration of switches are showed in the figure above, with e1/0/1, e1/0/10 and e1/0/15 all belonging to VLAN 100. The requirement is that, after port isolation is enabled on switch S1, e1/0/1 and e1/0/10 on switch S1 can not communicate with each other, while both of them can communicate with the uplink port e1/0/15. That is, the communication between any pair of downlink ports is disabled while that between any downlink port and a specified uplink port is normal. The uplink port can communicate with any port normally.

The configuration of S1:

Switch(config)#isolate-port group test

Switch(config)#isolate-port group test switchport interface ethernet 1/0/1;1/0/10

2.3 Port Loopback Detection

2.3.1 Introduction to Port Loopback Detection Function

With the development of switches, more and more users begin to access the network through Ethernet switches. In enterprise network, users access the network through layer-2 switches, which means urgent demands for both internet and the internal layer 2 Interworking. When layer 2 Interworking is required, the messages will be forwarded through MAC addressing the accuracy of which is the key to a correct Interworking between users. In layer 2 switching, the

messages are forwarded through MAC addressing. Layer 2 devices learn MAC addresses via learning source MAC address, that is, when the port receives a message from an unknown source MAC address, it will add this MAC to the receive port, so that the following messages with a destination of this MAC can be forwarded directly, which also means learn the MAC address once and for all to forward messages.

When a new source MAC is already learnt by the layer 2 device, only with a different source port, the original source port will be modified to the new one, which means to correspond the original MAC address with the new port. As a result, if there is any loopback existing in the link, all MAC addresses within the whole layer 2 network will be corresponded with the port where the loopback appears (usually the MAC address will be frequently shifted from one port to another), causing the layer 2 network collapsed. That is why it is a necessity to check port loopbacks in the network. When a loopback is detected, the detecting device should send alarms to the network management system, ensuring the network manager is able to discover, locate and solve the problem in the network and protect users from a long-lasting disconnected network.

Since detecting loopbacks can make dynamic judgment of the existence of loopbacks in the link and tell whether it has gone, the devices supporting port control (such as port isolation and port MAC address learning control) can maintain that automatically, which will not only reduce the burden of network managers but also response time, minimizing the effect caused loopbacks to the network.

2.3.2 Port Loopback Detection Function Configuration

Task List

1. Configure the time interval of loopback detection
2. Enable the function of port loopback detection
3. Configure the control method of port loopback detection
4. Display and debug the relevant information of port loopback detection
5. Configure the loopback-detection control mode (automatic recovery enabled or not)

1. Configure the time interval of loopback detection

Command	Explanation
Global Mode	
loopback-detection interval-time *<loopback> <no-loopback>* **no loopback-detection interval-time**	Configure the time interval of loopback detection.

2. Enable the function of port loopback detection

Command	Explanation
Port Mode	

loopback-detection specified-vlan <vlan-*list*> **no loopback-detection specified-vlan <vlan-*list*>**	Enable and disable the function of port loopback detection.

3. Configure the control method of port loopback detection

Command	Explanation
Port Mode	
loopback-detection control {shutdown |block| learning} **no loopback-detection control**	Enable and disable the function of port loopback detection control.

4. Display and debug the relevant information of port loopback detection

Command	Explanation
Admin Mode	
debug loopback-detection **no debug loopback-detection**	Enable the debug information of the function module of port loopback detection. The no operation of this command will disable the debug information.
show loopback-detection [interface <*interface-list*>]	Display the state and result of the loopback detection of all ports, if no parameter is provided; otherwise, display the state and result of the corresponding ports.

5. Configure the loopback-detection control mode (automatic recovery enabled or not)

Command	Explanation
Global Mode	
loopback-detection control-recovery timeout <0-3600>	Configure the loopback-detection control mode (automatic recovery enabled or not) or recovery time.

2.3.3 Port Loopback Detection Function Example



Figure 2-3 Typical example of port loopback detection

As shown in the above configuration, the switch will detect the existence of loopbacks in the network topology. After enabling the function of loopback detection on the port connecting the switch with the outside network, the switch will notify the connected network about the existence of a loopback, and control the port on the switch to guarantee the normal operation of the whole network.

The configuration task sequence of SWITCH:

Switch(config)#loopback-detection interval-time 35 15

Switch(config)#interface ethernet 1/0/1

Switch(Config-If-Ethernet1/0/1)#loopback-detection special-vlan 1-3

Switch(Config-If-Ethernet1/0/1)#loopback-detection control block

If adopting the control method of block, MSTP should be globally enabled. And the corresponding relation between the spanning tree instance and the VLAN should be configured.

Switch(config)#spanning-tree

Switch(config)#spanning-tree mst configuration

Switch(Config-Mstp-Region)#instance 1 vlan 1

Switch(Config-Mstp-Region)#instance 2 vlan 2

Switch(Config-Mstp-Region)#

2.3.4 Port Loopback Detection Troubleshooting

The function of port loopback detection is disabled by default and should only be enabled if required.

2.4 ULDP

2.4.1 Introduction to ULDP Function

Unidirectional link is a common error state of link in networks, especially in fiber links. Unidirectional link means that only one port of the link can receive messages from the other port, while the latter one can not receive messages from the former one. Since the physical layer of the link is connected and works normal, via the checking mechanism of the physical layer, communication problems between the devices can not be found. As shown in Graph, the problem in fiber connection can not be found through mechanisms in physical layer like automatic negotiation.

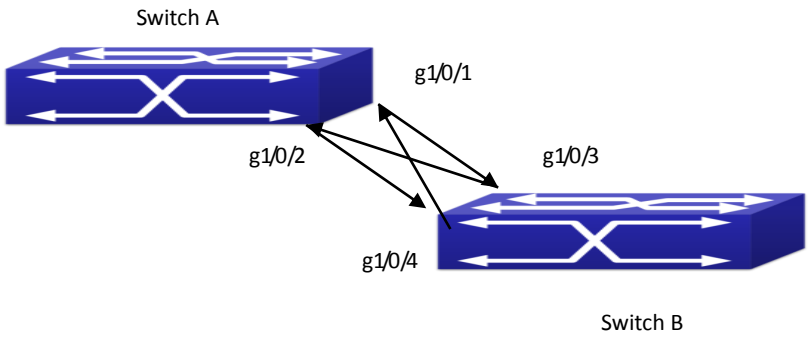


Figure 2-4 Fiber Cross Connection

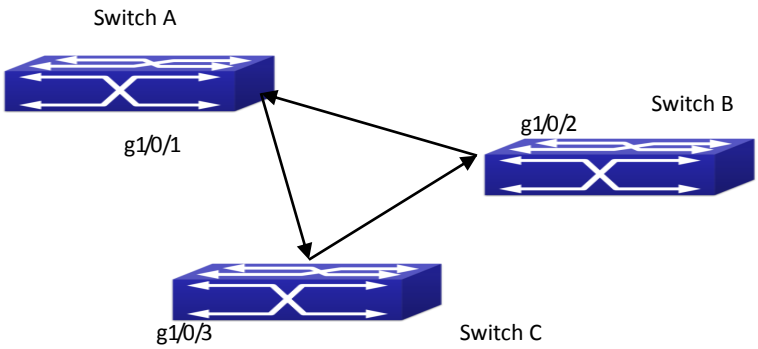


Figure 2-5 One End of Each Fiber Not Connected

This kind of problem often appears in the following situations: GBIC (Giga Bitrate Interface Converter) or interfaces have problems, software problems, hardware becomes unavailable or operates abnormally. Unidirectional link will cause a series of problems, such as spinning tree

topological loop, broadcast black hole.

ULDP (Unidirectional Link Detection Protocol) can help avoid disasters that could happen in the situations mentioned above. In a switch connected via fibers or copper Ethernet line (like ultra five-kind twisted pair), ULDP can monitor the link state of physical links. Whenever a unidirectional link is discovered, it will send warnings to users and can disable the port automatically or manually according to users' configuration.

The ULDP of switches recognizes remote devices and check the correctness of link connections via interacting ULDP messages. When ULDP is enabled on a port, protocol state machine will be started, which means different types of messages will be sent at different states of the state machine to check the connection state of the link by exchanging information with remote devices. ULDP can dynamically study the interval at which the remote device sends notification messages and adjust the local TTL (time to live) according to that interval. Besides, ULDP provides the reset mechanism, when the port is disabled by ULDP, it can check again through reset mechanism. The time intervals of notification messages and reset in ULDP can be configured by users, so that ULDP can respond faster to connection errors in different network environments.

The premise of ULDP working normally is that link works in duplex mode, which means ULDP is enabled on both ends of the link, using the same method of authentication and password.

2.4.2 ULDP Configuration Task Sequence

1. Enable ULDP function globally
2. Enable ULDP function on a port
3. Configure aggressive mode globally
4. Configure aggressive mode on a port
5. Configure the method to shut down unidirectional link
6. Configure the interval of Hello messages
7. Configure the interval of Recovery
8. Reset the port shut down by ULDP
9. Display and debug the relative information of ULDP

1. Enable ULDP function globally

Command	Explanation
Global configuration mode	
uldp enable **uldp disable**	Globally enable or disable ULDP function.

2. Enable ULDP function on a port

Command	Explanation
Port configuration mode	
uldp enable **uldp disable**	Enable or disable ULDP function on a port.

3. Configure aggressive mode globally

Command	Explanation
Global configuration mode	
uldp aggressive-mode **no uldp aggressive-mode**	Set the global working mode.

4. Configure aggressive mode on a port

Command	Explanation
Port configuration mode	
uldp aggressive-mode **no uldp aggressive-mode**	Set the working mode of the port.

5. Configure the method to shut down unidirectional link

Command	Explanation
Global configuration mode	
uldp manual-shutdown **no uldp manual-shutdown**	Configure the method to shut down unidirectional link.

6. Configure the interval of Hello messages

Command	Explanation
Global configuration mode	
uldp hello-interval *<integer>* **no uldp hello-interval**	Configure the interval of Hello messages, ranging from 5 to 100 seconds. The value is 10 seconds by default.

7. Configure the interval of Recovery

Command	Explanation
Global configuration mode	
uldp recovery-time *<integer>* **no uldp recovery-time *<integer>***	Configure the interval of Recovery reset, ranging from 30 to 86400 seconds. The value is 0 second by default.

8. Reset the port shut down by ULDP

Command	Explanation
Global configuration mode or port configuration mode	
uldp reset	Reset all ports in global configuration mode; Reset the specified port in port configuration mode.

9. Display and debug the relative information of ULDP

Command	Explanation

Admin mode	
show uldp [interface ethernet IFNAME]	Display ULDP information. No parameter means to display global ULDP information. The parameter specifying a port will display global information and the neighbor information of the port.
debug uldp fsm interface ethernet *<IFname>* **no debug uldp fsm interface ethernet *<IFname>***	Enable or disable the debug switch of the state machine transition information on the specified port.
debug uldp error **no debug uldp error**	Enable or disable the debug switch of error information.
debug uldp event **no debug uldp event**	Enable or disable the debug switch of event information.
debug uldp packet {receive|send} **no debug uldp packet {receive|send}**	Enable or disable the type of messages can be received and sent on all ports.
debug uldp {hello|probe|echo| unidir|all} [receive|send] interface ethernet *<IFname>* **no debug uldp {hello|probe|echo| unidir|all} [receive|send] interface ethernet *<IFname>***	Enable or disable the content detail of a particular type of messages can be received and sent on the specified port.

2.4.3 ULDP Function Typical Examples

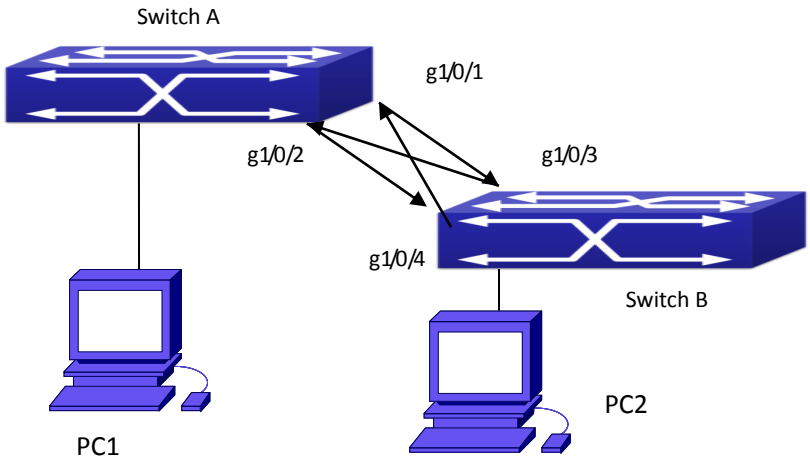


Figure 2-6 Fiber Cross Connection

In the network topology in Graph, port g1/0/1 and port g1/0/2 of SWITCH A as well as port g1/0/3 and port g1/0/4 of SWITCH B are all fiber ports. And the connection is cross connection. The physical layer is connected and works normally, but the data link layer is abnormal. ULDP can discover and disable this kind of error state of link. The final result is that port g1/0/1, g1/0/2 of

SWITCH A and port g1/0/3, g1/0/4 of SWITCH B are all shut down by ULDP. Only when the connection is correct, can the ports work normally (won't be shut down).

Switch A configuration sequence:

SwitchA(config)#uldp enable

SwitchA(config)#interface ethernet 1/0/1

SwitchA(Config-If-Ethernet1/0/1)#uldp enable

SwitchA(Config-If-Ethernet1/0/1)#exit

SwitchA(config)#interface ethernet 1/0/2

SwitchA(Config-If-Ethernet1/0/2)#uldp enable

Switch B configuration sequence:

SwitchB(config)#uldp enable

SwitchB(config)#interface ethernet1/0/3

SwitchB(Config-If-Ethernet1/0/3)#uldp enable

SwitchB(Config-If-Ethernet1/0/3)#exit

SwitchB(config)#interface ethernet 1/0/4

SwitchB(Config-If-Ethernet1/0/4)#uldp enable

As a result, port g1/0/1, g1/0/2 of SWITCH A are all shut down by ULDP, and there is notification information on the CRT terminal of PC1.

%Oct 29 11:09:50 2007 A unidirectional link is detected! Port Ethernet1/0/1 need to be shutted down!

%Oct 29 11:09:50 2007 Unidirectional port Ethernet1/0/1 shut down!

%Oct 29 11:09:50 2007 A unidirectional link is detected! Port Ethernet1/0/2 need to be shutted down!

%Oct 29 11:09:50 2007 Unidirectional port Ethernet1/0/2 shutted down!

Port g1/0/3, and port g1/0/4 of SWITCH B are all shut down by ULDP, and there is notification information on the CRT terminal of PC2.

%Oct 29 11:09:50 2007 A unidirectional link is detected! Port Ethernet1/0/3 need to be shutted down!

%Oct 29 11:09:50 2007 Unidirectional port Ethernet1/0/3 shutted down!

%Oct 29 11:09:50 2007 A unidirectional link is detected! Port Ethernet1/0/4 need to be shutted down!

%Oct 29 11:09:50 2007 Unidirectional port Ethernet1/0/4 shutted down!

2.4.4 ULDP Troubleshooting

Configuration Notice:

- ☞ In order to ensure that ULDP can discover that the one of fiber ports has not connected or the ports are incorrectly cross connected, the ports have to work in duplex mode and have the same rate.
- ☞ If the automatic negotiation mechanism of the fiber ports with one port misconnected decides the working mode and rate of the ports, ULDP won't take effect no matter enabled or not. In such situation, the port is considered as "Down".
- ☞ In order to make sure that neighbors can be correctly created and unidirectional links can be

correctly discovered, it is required that both end of the link should enable ULDP, using the same authentication method and password. At present, no password is needed on both ends.

- ☞ The hello interval of sending hello messages can be changed (it is10 seconds by default and ranges from 5 to 100 seconds) so that ULDP can respond faster to connection errors of links in different network environments. But this interval should be less than 1⁄3 of the STP convergence time. If the interval is too long, a STP loop will be generated before ULDP discovers and shuts down the unidirectional connection port. If the interval is too short, the network burden on the port will be increased, which means a reduced bandwidth.
- ☞ ULDP does not handle any LACP event. It treats every link of TRUNK group (like Port-channel, TRUNK ports) as independent, and handles each of them respectively.
- ☞ ULDP does not compact with similar protocols of other vendors, which means users can not use ULDP on one end and use other similar protocols on the other end.
- ☞ ULDP function is disabled by default. After globally enabling ULDP function, the debug switch can be enabled simultaneously to check the debug information. There are several DEBUG commands provided to print debug information, such as information of events, state machine, errors and messages. Different types of message information can also be printed according to different parameters.
- ☞ The Recovery timer is disabled by default and will only be enabled when the users have configured recovery time (30-86400 seconds).
- ☞ Reset command and reset mechanism can only reset the ports automatically shut down by ULDP. The ports shut down manually by users or by other modules won't be reset by ULDP.

2.5 LLDP

2.5.1 Introduction to LLDP Function

Link Layer Discovery Protocol (LLDP) is a new protocol defined in 802.1ab. It enables neighbor devices to send notices of their own state to other devices, and enables all ports of every device to store information about them. If necessary, the ports can also send update information to the neighbor devices directly connected to them, and those neighbor devices will store the information in standard SNMP MIBs. The network management system can check the layer-two connection state from MIB. LLDP won't configure or control network elements or flows, but only report the configuration of layer-two. Another content of 802.1ab is to utilizing the information provided by LLDP to find the conflicts in layer-two. IEEE now uses the existing physical topology, interfaces and Entity MIBs of IETF.

To simplify, LLDP is a neighbor discovery protocol. It defines a standard method for Ethernet

devices, such as switches, routers and WLAN access points, to enable them to notify their existence to other nodes in the network and store the discovery information of all neighbor devices. For example, the detail information of the device configuration and discovery can both use this protocol to advertise.

In specific, LLDP defines a general advertisement information set, a transportation advertisement protocol and a method to store the received advertisement information. The device to advertise its own information can put multiple pieces of advertisement information in one LAN data packet to transport. The type of transportation is the type length value (TLV) field. All devices supporting LLDP have to support device ID and port ID advertisement, but it is assumed that, most devices should also support system name, system description and system performance advertisement. System name and system description advertisement can also provide useful information for collecting network flow data. System description advertisement can include data such as the full name of the advertising device, hardware type of system, the version information of software operation system and so on.

802.1AB Link Layer Discovery Protocol will make searching the problems in an enterprise network an easier process and can strengthen the ability of network management tools to discover and maintain accurate network topology structure.

Many kinds of network management software use “Automated Discovery” function to trace the change and condition of topology, but most of them can reach layer-three and classify the devices into all IP subnets at best. This kind of data are very primitive, only referring to basic events like the adding and removing of relative devices instead of details about where and how these devices operate with the network.

Layer 2 discovery covers information like which devices have which ports, which switches connect to other devices and so on, it can also display the routs between clients, switches, routers, application servers and network servers. Such details will be very meaningful for schedule and investigate the source of network failure.

LLDP will be a very useful management tool, providing accurate information about network mirroring, flow data and searching network problems.

2.5.2 LLDP Function Configuration Task Sequence

1. Globally enable LLDP function
2. Configure the port-based LLDP function switch
3. Configure the operating state of port LLDP
4. Configure the intervals of LLDP updating messages
5. Configure the aging time multiplier of LLDP messages
6. Configure the sending delay of updating messages
7. Configure the intervals of sending Trap messages
8. Configure to enable the Trap function of the port
9. Configure the optional information-sending attribute of the port
10. Configure the size of space to store Remote Table of the port
11. Configure the type of operation when the Remote Table of the port is full
12. Display and debug the relative information of LLDP

1. Globally enable LLDP function

Command	Explanation
Global Mode	
lldp enable **lldp disable**	Globally enable or disable LLDP function.

2. Configure the port-base LLDP function switch

Command	Explanation
Port Mode	
lldp enable **lldp disable**	Configure the port-base LLDP function switch.

3. Configure the operating state of port LLDP

Command	Explanation
Port Mode	
lldp mode (send|receive|both|disable)	Configure the operating state of port LLDP.

4. Configure the intervals of LLDP updating messages

Command	Explanation
Global Mode	
lldp tx-interval ***<integer>*** **no lldp tx-interval**	Configure the intervals of LLDP updating messages as the specified value or default value.

5. Configure the aging time multiplier of LLDP messages

Command	Explanation
Global Mode	
lldp msgTxHold ***<value>*** **no lldp msgTxHold**	Configure the aging time multiplier of LLDP messages as the specified value or default value.

6. Configure the sending delay of updating messages

Command	Explanation
Global Mode	
lldp transmit delay ***<seconds>*** **no lldp transmit delay**	Configure the sending delay of updating messages as the specified value or default value.

7. Configure the intervals of sending Trap messages

Command	Explanation
Global Mode	

lldp notification interval <*seconds*> **no lldp notification interval**	Configure the intervals of sending Trap messages as the specified value or default value.

8. Configure to enable the Trap function of the port

Command	Explanation
Port Configuration Mode	
lldp trap <*enable*|*disable*>	Enable or disable the Trap function of the port.

9. Configure the optional information-sending attribute of the port

Command	Explanation
Port Configuration Mode	
lldp transmit optional tlv [portDesc] [sysName] [sysDesc] [sysCap] **no lldp transmit optional tlv**	Configure the optional information-sending attribute of the port as the option value of default values.

10. Configure the size of space to store Remote Table of the port

Command	Explanation
Port Configuration Mode	
lldp neighbors max-num < *value* > **no lldp neighbors max-num**	Configure the size of space to store Remote Table of the port as the specified value or default value.

11. Configure the type of operation when the Remote Table of the port is full

Command	Explanation
Port Configuration Mode	
lldp tooManyNeighbors {discard | delete}	Configure the type of operation when the Remote Table of the port is full.

12. Display and debug the relative information of LLDP

Command	Explanation
Admin, Global Mode	
show lldp	Display the current LLDP configuration information.
show lldp interface ethernet <*IFNAME*>	Display the LLDP configuration information of the current port.
show lldp traffic	Display the information of all kinds of counters.
show lldp neighbors interface ethernet < *IFNAME* >	Display the information of LLDP neighbors of the current port.

show debugging lldp	Display all ports with LLDP debug enabled.
Admin Mode	
debug lldp **no debug lldp**	Enable or disable the DEBUG switch.
debug lldp packets interface ethernet *<IFNAME>* **no debug lldp packets interface ethernet *<IFNAME>***	Enable or disable the DEBUG packet-receiving and sending function in port or global mode.
Port configuration mode	
clear lldp remote-table	Clear Remote-table of the port.

2.5.3 LLDP Function Typical Example

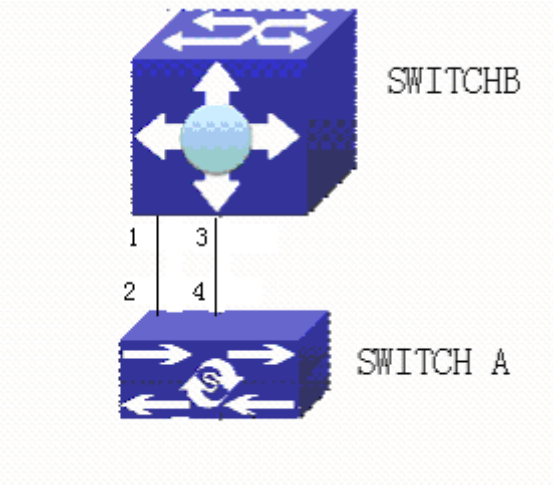


Figure 2-7 LLDP Function Typical Configuration Example

In the network topology graph above, the port 1,3 of SWITCH B are connected to port 2,4 of SWITCH A. Port 1 of SWITCH B is configured to message-receiving-only mode, Option TLV of port 4 of SWITCH A is configured as portDes and SysCap.

SWITCH A configuration task sequence:

SwitchA(config)# lldp enable

SwitchA(config)#interface ethernet 1/0/4

SwitchA(Config-If-Ethernet1/0/4)#lldp transmit optional tlv portDesc sysCap

SwitchA(Config-If-Ethernet1/0/4)exit

SWITCH B configuration task sequence:

SwitchB(config)#lldp enable

SwitchB(config)#interface ethernet1/0/1

SwitchB(Config-If-Ethernet1/0/1)#lldp mode receive

SwitchB(Config-If-Ethernet1/0/1)#exit

2.5.4 LLDP Function Troubleshooting

- ☞ LLDP function is disabled by default. After enabling the global switch of LLDP, users can enable the debug switch "**debug lldp**" simultaneously to check debug information.
- ☞ Using "show" function of LLDP function can display the configuration information in global or port configuration mode.

2.6 LLDP-MED

2.6.1 Introduction to LLDP-MED

LLDP-MED (Link Layer Discovery Protocol-Media Endpoint Discovery) based on 802.1AB LLDP (Link Layer Discovery Protocol) of IEEE. LLDP provides a standard link layer discovery mode, it sends local device information (including its major capability, management IP address, device ID and port ID) as TLV (type/length/value) triplets in LLDPDU (Link Layer Discovery Protocol Data Unit) to the direct connection neighbors. The device information received by the neighbors will be stored with a standard management information base (MIB). This allows a network management system to quickly detect and identify the communication status of the link.

In 802.1AB LLDP, there is no transmission and management about the voice device information. To deploy and manage voice device expediently, LLDP-MED TLVs provide multiple information, such as PoE (Power over Ethernet), network policy, and the location information of the emergent telephone service.

2.6.2 LLDP-MED Configuration Task Sequence

1. Basic LLDP-MED configuration

Command	Explanation
Port mode	
lldp transmit med tlv all **no lldp transmit med tlv all**	Configure the specified port to send all LLDP-MED TLVs. The no command disables the function.
lldp transmit med tlv capability **no lldp transmit med tlv capability**	Configure the specified port to send LLDP-MED Capability TLV. The no command disables the capability.
lldp transmit med tlv networkPolicy **no lldp transmit med tlv networkPolicy**	Configure the specified port to send LLDP-MED Network Policy TLV. The no command disables the capability.

lldp transmit med tlv extendPoe **no lldp transmit med tlv extendPoe**	Configure the specified port to send LLDP-MED Extended Power-Via-MDI TLV. The no command disables the capability.
lldp transmit med tlv location **no lldp transmit med tlv location**	Configure the specified port to send LLDP-MED Location Identification TLV. The no command disables the capability.
lldp transmit med tlv inventory **no lldp transmit med tlv inventory**	Configure the port to send LLDP-MED Inventory Management TLVs. The no command disables the capability.
network policy {voice | voice-signaling | guest-voice | guest-voice-signaling | softphone-voice | video-conferencing | streaming-video | video-signaling} [status {enable | disable}] [tag {tagged | untagged}] [vid {*<vlan-id>* | dot1p}] [cos <*cos-value*>] [dscp <*dscp-value*>] **no network policy {voice | voice-signaling | guest-voice | guest-voice-signaling | softphone-voice | video-conferencing | streaming- video | video-signaling}**	Configure network policy of the port, including VLAN ID, the supported application (such as voice and video), the application priority and the used policy, and so on.
civic location {dhcp server | switch | endpointDev} *<country-code>* **no civic location**	Configure device type and country code of the location with Civic Address LCI format and enter Civic Address LCI address mode. The no command cancels all configurations of the location with Civic Address LCI format.
ecs location *<tel-number>* **no ecs location**	Configure the location with ECS ELIN format on the port, the no command cancels the configured location.
lldp med trap {enable | disable}	Enable or disable LLDP-MED trap for the specified port.
Civic Address LCI address mode	
{description-language | province-state | city | county | street | locationNum | location | floor | room | postal | otherInfo} *<address>* **no {description-language | province-state | city | county | street | locationNum | location | floor | room | postal**	Configure the detailed address after enter Civic Address LCI address mode of the port.

| otherInfo}	
Global mode	
lldp med fast count <*value*> **no lldp med fast count**	When the fast LLDP-MED startup mechanism is enabled, it needs to fast send the LLDP packets with LLDP-MED TLV, this command is used to set the value of the fast sending packets, the no command restores the default value.
Admin mode	
show lldp	Show the configuration of the global LLDP and LLDP-MED.
show lldp [interface ethernet <*IFNAME*>]	Show the configuration of LLDP and LLDP-MED on the current port.
show lldp neighbors [interface ethernet <*IFNAME*>]	Show LLDP and LLDP-MED configuration of the neighbors.
show lldp traffic	Show the statistics of the sent and received packets of port LLDP and LLDP-MED.

2.6.3 LLDP-MED Example

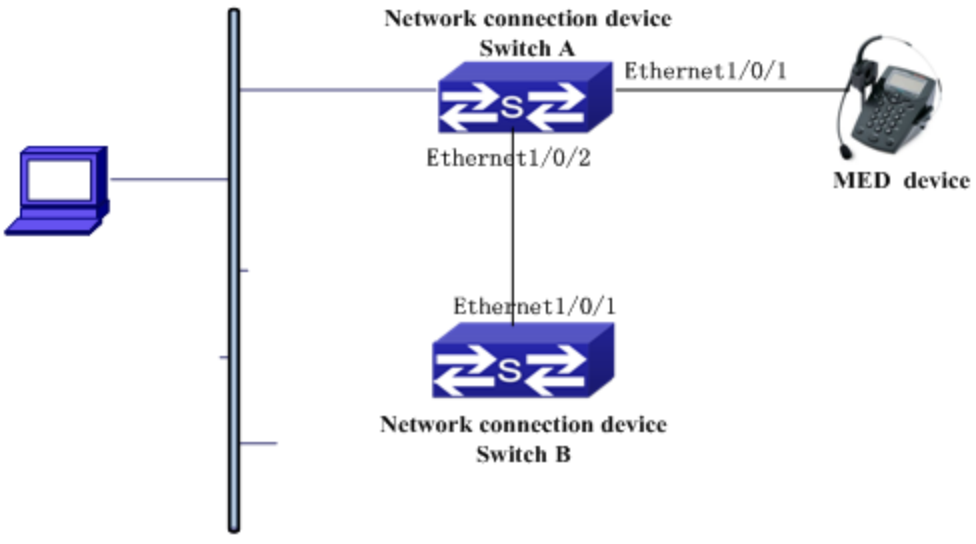


Figure 2-8 Basic LLDP-MED configuration topology

1) Configure Switch A

SwitchA(config)#interface ethernet1/0/1

SwitchA (Config-If-Ethernet1/0/1)# lldp enable

SwitchA (Config-If-Ethernet1/0/1)# lldp mode both（this configuration can be omitted, the default

mode is RxTx）
SwitchA (Config-If-Ethernet1/0/1)# lldp transmit med tlv capability
SwitchA (Config-If-Ethernet1/0/1)# lldp transmit med tlv network policy
SwitchA (Config-If-Ethernet1/0/1)# lldp transmit med tlv inventory
SwitchB (Config-If-Ethernet1/0/1)# network policy voice tag tagged vid 10 cos 5 dscp 15
SwitchA (Config-If-Ethernet1/0/1)# exit
SwitchA (config)#interface ethernet1/0/2
SwitchA (Config-If-Ethernet1/0/2)# lldp enable
SwitchA (Config-If-Ethernet1/0/2)# lldp mode both
2) Configure Switch B
SwitchB (config)#interface ethernet1/0/1
SwitchB(Config-If-Ethernet1/0/1)# lldp enable
SwitchB (Config-If-Ethernet1/0/1)# lldp mode both
SwitchB (Config-If-Ethernet1/0/1)# lldp transmit med tlv capability
SwitchB (Config-If-Ethernet1/0/1)# lldp transmit med tlv network policy
SwitchB (Config-If-Ethernet1/0/1)# lldp transmit med tlv inventory
SwitchB (Config-If-Ethernet1/0/1)# network policy voice tag tagged vid 10 cos 4
3) Verify the configuration
Show the global status and interface status on Switch A.
SwitchA# show lldp neighbors interface ethernet 1/0/1
Port name : Ethernet1/0/1
Port Remote Counter : 1
TimeMark :20
ChassisIdSubtype :4
ChassisId :00-03-0f-00-00-02
PortIdSubtype :Local
PortId :1
PortDesc :****
SysName :****
SysDesc :*****

SysCapSupported :4
SysCapEnabled :4

LLDP MED Information :
MED Codes:
(CAP)Capabilities, (NP) Network Policy
(LI) Location Identification, (PSE)Power Source Entity
(PD) Power Device, (IN) Inventory
MED Capabilities:CAP,NP,PD,IN
MED Device Type: Endpoint Class III
Media Policy Type :Voice
Media Policy :Tagged
Media Policy Vlan id :10

Media Policy Priority :3
Media Policy Dscp :5
Power Type : PD
Power Source :Primary power source
Power Priority :low
Power Value :15.4 (Watts)
Hardware Revision:
Firmware Revision:4.0.1
Software Revision:6.2.30.0
Serial Number:
Manufacturer Name:****
Model Name:Unknown
Assert ID:Unknown
IEEE 802.3 Information :
auto-negotiation support: Supported
auto-negotiation support: Not Enabled
PMD auto-negotiation advertised capability: 1
operational MAU type: 1
SwitchA# show lldp neighbors interface ethernet 1/0/2
Port name : interface ethernet 1/0/2
Port Remote Counter： 1
Neighbor Index: 1
Port name : Ethernet1/0/2
Port Remote Counter : 1
TimeMark :20
ChassisIdSubtype :4
ChassisId :00-03-0f-00-00-02
PortIdSubtype :Local
PortId :1
PortDesc :Ethernet1/0/1
SysName :****
SysDesc :*****

SysCapSupported :4
SysCapEnabled :4

Explanation:
1) Both Ethernet2 of switch A and Ethernet1 of switch B are the ports of network connection device, they will not send LLDP packets with MED TLV information forwardly. Although configure Ethernet1 of switch B to send MED TLV information, it will not send the related MED information, that results the corresponding Remote table without the related MDE information on Ethernet2 of switch A.
2) LLDP-MED device is able to send LLDP packets with MED TLV forwardly, so the corresponding Remote table with LLDP MED information on Ethernet1 of switch A.

2.6.4 LLDP-MED Troubleshooting

If problems occur when configuring LLDP-MED, please check whether the problem is caused by the following reasons:

- Check whether the global LLDP is enabled.
- Only network connection device received LLDP packets with LLDP-MED TLV from the near MED device, it sends LLDP-MED TLV. If network connection device configured the command for sending LLDP-MED TLV, the packets also without LLDP-MED TLV sent by the port, that means no MED information is received and the port does not enable the function for sending LLDP-MED information.
- If neighbor device has sent LLDP-MED information to network connection device, but there is no LLDP-MED information by checking **show lldp neighbors** command, that means LLDP-MED information sent by neighbor is error.

2.7 Port Channel

2.7.1 Introduction to Port Channel

To understand Port Channel, Port Group should be introduced first. Port Group is a group of physical ports in the configuration level; only physical ports in the Port Group can take part in link aggregation and become a member port of a Port Channel. Logically, Port Group is not a port but a port sequence. Under certain conditions, physical ports in a Port Group perform port aggregation to form a Port Channel that has all the properties of a logical port, therefore it becomes an independent logical port. Port aggregation is a process of logical abstraction to abstract a set of ports (port sequence) with the same properties to a logical port. Port Channel is a collection of physical ports and used logically as one physical port. Port Channel can be used as a normal port by the user, and can not only add network's bandwidth, but also provide link backup. Port aggregation is usually used when the switch is connected to routers, PCs or other switches.

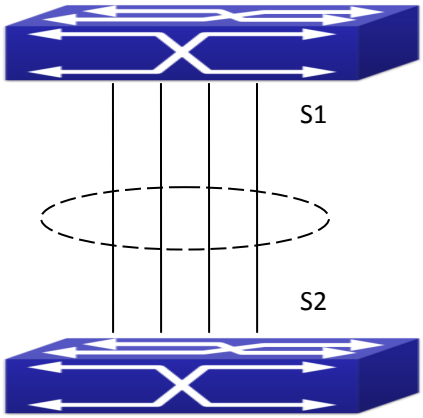


Figure 2-9 Port aggregation

As shown in the above, S1 is aggregated to a Port Channel, the bandwidth of this Port Channel is the total of all the four ports. If traffic from S1 needs to be transferred to S2 through the Port Channel, traffic allocation calculation will be performed based on the source MAC address and the lowest bit of target MAC address. The calculation result will decide which port to convey the traffic. If a port in Port Channel fails, the other ports will undertake traffic of that port through a traffic allocation algorithm. This algorithm is carried out by the hardware.

Switch offers two methods for configuring port aggregation: manual Port Channel creation and LACP (Link Aggregation Control Protocol) dynamic Port Channel creation. Port aggregation can only be performed on ports in full-duplex mode.

For Port Channel to work properly, member ports of the Port Channel must have the same properties as follows:

- All ports are in full-duplex mode.
- All Ports are of the same speed.
- All ports are Access ports and belong to the same VLAN or are all TRUNK ports, or are all Hybrid ports.
- If the ports are all TRUNK ports or Hybrid ports, then their “Allowed VLAN” and “Native VLAN” property should also be the same.

If Port Channel is configured manually or dynamically on switch, the system will automatically set the port with the smallest number to be Master Port of the Port Channel. If the spanning tree function is enabled in the switch, the spanning tree protocol will regard Port Channel as a logical port and send BPDU frames via the master port.

Port aggregation is closely related with switch hardware. Switch allow physical port aggregation of any two switches, maximum 128 groups and 8 ports in each port group are supported.

Once ports are aggregated, they can be used as a normal port. Switch have a built-in aggregation interface configuration mode, the user can perform related configuration in this mode just like in the VLAN and physical interface configuration mode.

2.7.2 Brief Introduction to LACP

LACP (Link Aggregation Control Protocol) is a kind of protocol based on IEEE802.3ad standard to implement the link dynamic aggregation. LACP protocol uses LACPDU (Link Aggregation Control Protocol Data Unit) to exchange the information with the other end.

After LACP protocol of the port is enabled, this port will send LACPDU to the other end to notify the system priority, the MAC address of the system, the priority of the port, the port ID and the operation Key. After the other end receives the information, the information is compared with the saving information of other ports to select the port which can be aggregated, accordingly, both sides can reach an agreement about the ports join or exit the dynamic aggregation group.

The operation Key is created by LACP protocol according to the combination of configuration (speed, duplex, basic configuration, management Key) of the ports to be aggregated.

After the dynamic aggregation port enables LACP protocol, the management Key is 0 by default. After the static aggregation port enables LACP, the management Key of the port is the same with the ID of the aggregation group.

For the dynamic aggregation group, the members of the same group have the same operation Key, for the static aggregation group, the ports of Active have the same operation Key.

The port aggregation is that multi-ports are aggregated to form an aggregation group, so as to implement the out/in load balance in each member port of the aggregation group and provides the better reliability.

2.7.2.1 Static LACP Aggregation

Static LACP aggregation is enforced by users configuration, and do not enable LACP protocol. When configuring static LACP aggregation, use “on” mode to force the port to enter the aggregation group.

2.7.2.2 Dynamic LACP Aggregation

1. The summary of the dynamic LACP aggregation

Dynamic LACP aggregation is an aggregation created/deleted by the system automatically, it does not allow the user to add or delete the member ports of the dynamic LACP aggregation. The ports which have the same attribute of speed and duplex, are connected to the same device, have the same basic configuration, can be dynamically aggregated together. Even if only one port can create the dynamic aggregation, that is the single port aggregation. In the dynamic aggregation, LACP protocol of the port is at the enable state.

2. The port state of the dynamic aggregation group

In dynamic aggregation group, the ports have two states: selected or standby. Both selected ports and standby ports can receive and send LACP protocol, but standby ports can not forward the data packets.

Because the limitation of the max port number in the aggregation group, if the current number of the member ports exceeds the limitation of the max port number, then the system of this end will negotiates with the other end to decide the port state according to the port ID. The negotiation steps are as follows:

Compare ID of the devices (the priority of the system + the MAC address of the system). First,

compare the priority of the systems, if they are same, then compare the MAC address of the systems. The end with a small device ID has the high priority.

Compare the ID of the ports (the priority of the port + the ID of the port). For each port in the side of the device which has the high device priority, first, compare the priority of the ports, if the priorities are same, then compare the ID of the ports. The port with a small port ID is selected, and the others become the standby ports.

In an aggregation group, the port which has the smallest port ID and is at the selected state will be the master port, the other ports at the selected state will be the member port.

2.7.3 Introduction to Load balance

The current visits and data flow of the network are increasing; the processing capability and calculated strength are both increasing. If the large amount of the data flow is transmitted from one physical port of the switch at the same, it will cause the network congestion. If there are many physical ports of the switch, it will cause the ports wasting. So there is a method which can expand the network device and server bandwidth, increase the throughout, improve the network flexibility and strengthen the data processing, it is Load Balance.

2.7.4 Port Channel Configuration Task List

1. Create a port group in Global Mode
2. Add ports to the specified group from the Port Mode of respective ports

3. Enter the load-balance enhanced profile mode
4. Configure the enhanced load balance template
5. Enter port-channel configuration mode
6. Set load-balance method for port-group
7. Set the system priority of LACP protocol
8. Set the port priority of the current port in LACP protocol
9. Set the timeout mode of the current port in LACP protocol

1. Creating a port group

Command	Explanation
Global Mode	
port-group <*port-group-number*> **no port-group <*port-group-number*>**	Create or delete a port group.

2. Add physical ports to the port group

Command	Explanation
Port Mode	

port-group *<port-group-number>* mode {active | passive | on} **no port-group**	Add the ports to the port group and set their mode.

3. Enter the load-balance enhanced profile mode

Command	Explanation
Global Mode	
load-balance enhanced profile	Enter the load-balance enhanced profile mode.

4. Configure the enhanced load balance template

Command	Explanation
Load-balance Enhanced Profile Mode	
l2 field [dst-mac] [ingress-port] [l2-protocol] [src-mac] [vlan] **no l2 field**	This command is used to configure the load-balance enhanced l2 packets field. The no command recovers to be the default configuration that means all the fields are configured.
l2 mpls field l2payload [dst-mac] [l2-protocol] [src-mac] [vlan] **no l2 mpls field**	This command is used to configure the load-balance enhanced l2 mpls field. The no command recovers to be the default configuration that means all the fields are configured.
l2 mpls field l3payload [dst-ip] [l4-dst-port] [l4-src-port] [nh-or-ipprotocol][src-ip] [vlan] **no l2 mpls field**	This command is used to configure the load-balance enhanced l2 mpls field. The no command recovers to be the default configuration that means all the fields are configured.
ipv4 field [dst-ip] [ingress-port] [l4-dst-port] [l4-src-port] [protocol] [src-ip][vlan] **no ipv4 field**	This command is used to configure the load-balance enhanced ipv4 packets field. The no command recovers to be the default configuration that means all the fields are configured.
ipv6 field [dst-ip] [ingress-port] [l4-dst-port] [l4-src-port] [protocol] [src-ip][vlan] **no ipv6 field**	This command is used to configure the load-balance enhanced ipv6 packets field. The no command recovers to be the default configuration that means all the fields are configured.
l3 mpls field [dst-ip] [l4-dst-port] [l4-src-port] [protocol] [src-ip] [vlan] **no l3** mpls **field**	This command is used to configure the load-balance enhanced l3 mpls field. The no command recovers to be the default configuration that means all the fields are configured.

mpls tunnel field [2nd-label] [3rd-label] [dst-ip] [label-4msb] [src-ip] [top-label] no **mpls tunnel field**	This command is used to configure the load-balance enhanced mpls tunnel field. The no command recovers to be the default configuration that means all the fields are configured.
mim field l2payload [dst-mac] [l2-protocol] [src-mac] [vlan] **no mim field**	This command is used to configure the load-balance enhanced mim l2payload field. The no command recovers to be the default configuration that means all the fields are configured.
mim field l3payload [dst-ip] [l4-dst-port] [l4-src-port] [protocol] [src-ip] [vlan] **no mim** field	This command is used to configure the load-balance enhanced mim l3payload field. The no command recovers to be the default configuration that means all the fields are configured.
mim tunnel field [dst-mac] [lookup-id] [src-mac] **no mim tunnel field**	This command is used to configure the load-balance enhanced mim tunnel field. The no command recovers to be the default configuration that means all the fields are configured.
trill field l2payload [dst-mac] [l2-protocol] [src-mac] [vlan] **no trill field**	This command is used to configure the load-balance enhanced trill l2payload field. The no command recovers to be the default configuration that means all the fields are configured.
trill field l3payload [dst-ip] [l4-dst-port] [l4-src-port] [protocol] [src-ip] [vlan] **no trill** field	This command is used to configure the load-balance enhanced trill l3payload field. The no command recovers to be the default configuration that means all the fields are configured.
trill tunnel field l2payload [dst-mac] [egr-rbridge-name][ing-rbridge-name] [l2-protocol] [src-mac] [vlan] **no trill tunnel field**	This command is used to configure the load-balance enhanced trill tunnel field. The no command recovers to be the default configuration that means all the fields are configured.
trill tunnel field l3payload **no trill tunnel field**	This command is used to configure the load-balance enhanced trill tunnel l3payload field which matching the trill field l3payload. The no command recovers to be the default configuration that means all the fields are configured.
trill tunnel field outerl2 **no trill tunnel field**	This command is used to configure the load-balance enhanced trill tunnel outer l2 field which matching l2 field.

	The no command recovers to be the default configuration that means all the fields are configured.

5. Enter port-channel configuration mode.

Command	Explanation
Global Mode	
interface port-channel <port-channel-number>	Enter port-channel configuration mode.

6. Set load-balance method for port-group

Command	Explanation
Aggregation port configuration mode	
load-balance {dsp-ip | dst-mac | dst-src-ip | dst-src-mac | enhance-profile | src-ip | src-mac }	Set load-balance for port-group.

7. Set the system priority of LACP protocol

Command	Explanation
Global mode	
lacp system-priority <system-priority> **no lacp system-priority**	Set the system priority of LACP protocol, the no command restores the default value.

8. Set the port priority of the current port in LACP protocol

Command	Explanation
Port mode	
lacp port-priority <port-priority> **no lacp port-priority**	Set the port priority in LACP protocol. The no command restores the default value.

9. Set the timeout mode of the current port in LACP protocol

Command	Explanation
Port mode	
lacp timeout {short | long} **no lacp timeout**	Set the timeout mode in LACP protocol. The no command restores the default value.

2.7.5 Port Channel Examples

Scenario 1: Configuring Port Channel in LACP.

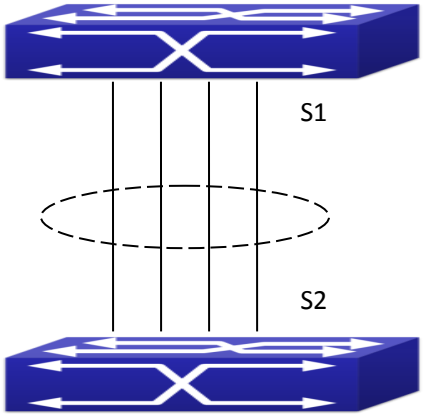


Figure 2-10 Configure Port Channel in LACP

The switches in the description below are all switch and as shown in the figure, ports 1, 2, 3, 4 of S1 are access ports and add them to group1 with active mode. Ports 6, 8, 9, 10 of S2 are access ports and add them to group2 with passive mode. All the ports should be connected with cables.

The configuration steps are listed below:

```
 Switch1#config
Switch1(config)#interface ethernet 1/1-4
Switch1(Config-If-Port-Range)#port-group 1 mode active
Switch1(Config-If-Port-Range)#exit
Switch1(config)#interface port-channel 1
Switch1(Config-If-Port-Channel1)#

Switch2#config
Switch2(config)#port-group 2
Switch2(config)#interface ethernet 1/6
Switch2(Config-If-Ethernet1/6)#port-group 2 mode passive
Switch2(Config-If-Ethernet1/6)#exit
Switch2(config)#interface ethernet 1/8-10
Switch2(Config-If-Port-Range)#port-group 2 mode passive
Switch2(Config-If-Port-Range)#exit
Switch2(config)#interface port-channel 2
Switch2(Config-If-Port-Channel2)#
```

Configuration result:

Shell prompts ports aggregated successfully after a while, now ports 1, 2, 3, 4 of S1 form an aggregated port named “Port-Channel1”, ports 6, 8, 9, 10 of S2 form an aggregated port named “Port-Channel2”; can be configured in their respective aggregated port mode.

Scenario 2: Configuring Port Channel in ON mode.

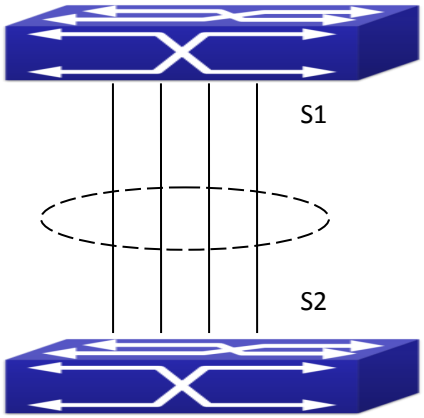


Figure 2-11 Configure Port Channel in ON mode

As shown in the figure, ports 1, 2, 3, 4 of S1 are access ports and add them to group1 with "on" mode. Ports 6, 8, 9, 10 of S2 are access ports and add them to group2 with "on" mode.

The configuration steps are listed below:

Switch1#config
Switch1(config)#interface ethernet 1/1
Switch1(Config-If-Ethernet1/1)#port-group 1 mode on
Switch1(Config-If-Ethernet1/1)#exit
Switch1(config)#interface ethernet 1/2
Switch1 (Config-If-Ethernet1/2)#port-group 1 mode on
Switch1 (Config-If-Ethernet1/2)#exit
Switch1 (config)#interface ethernet 1/3
Switch1 (Config-If-Ethernet1/3)#port-group 1 mode on
Switch1 (Config-If-Ethernet1/3)#exit
Switch1 (config)#interface ethernet 1/4
Switch1 (Config-If-Ethernet1/4)#port-group 1 mode on
Switch1 (Config-If-Ethernet1/4)#exit

Switch2#config
Switch2(config)#port-group 2
Switch2(config)#interface ethernet 1/6
Switch2 (Config-If-Ethernet1/6)#port-group 2 mode on
Switch2 (Config-If-Ethernet1/6)#exit
Switch2 (config)#interface ethernet 1/8-10
Switch2(Config-If-Port-Range)#port-group 2 mode on
Switch2(Config-If-Port-Range)#exit

Configuration result:

Add ports 1, 2, 3, 4 of S1 to port-group1 in order, and we can see a group in 'on' mode is completely joined forcedly, switch in other ends won't exchange LACP PDU to complete aggregation. Aggregation finishes immediately when the command to add port 1/2 to port-group 1 is entered, port 1 and port 2 aggregate to be port-channel 1, when port 1/3 joins port-group 1, port-channel 1 of port 1 and 2 are ungrouped and re-aggregate with port 3 to form port-channel 1, when port 1/4 joins port-group 1, port-channel 1 of port 1, 2 and 3 are ungrouped and re-aggregate with port 4 to form port-channel 1. (It should be noted that whenever a new port joins in an aggregated port group, the group will be ungrouped first and re-aggregated to form a new group.) Now all four ports in both S1 and S2 are aggregated in 'on' mode and become an aggregated port respectively.

2.7.6 Troubleshooting

2.7.6.1 Port Channel Troubleshooting

If problems occur when configuring port aggregation, please first check the following for causes.

- ☞ Ensure all ports in a port group have the same properties, i.e., whether they are in full-duplex mode, forced to the same speed, and have the same VLAN properties, etc. If inconsistency occurs, make corrections.
- ☞ Some commands cannot be used on a port in port-channel, such as arp, bandwidth, ip, ip-forward, etc.

2.7.6.2 Load Balance Troubleshooting

- ☞ If problems occur when configuring Load Balance, please first check the following for causes.
  - ☞ Check if the port-channel port aggregation is up by the command of show interface port channel.
- ☞ Check if the enhanced load balance template is configured correctly by the command of show load-balance enhanced-profile.

2.8 MTU

2.8.1 Introduction to MTU

So far the Jumbo (Jumbo Frame) has not reach a determined standard in the industry (including the format and length of the frame). Normally frames sized within 1519-12000 should be considered jumbo frame. Networks with jumbo frames will increase the speed of the whole network by 2% to 5%. Technically the Jumbo is just a lengthened frame sent and received by the switch. However considering the length of Jumbo frames, they will not be sent to CPU. We discard the Jumbo frames sent to CPU in the packet receiving process.

2.8.2 MTU Configuration Task Sequence

1. Configure enable MTU function

1. Configure enable MTU function

Command	Explanation
Global Mode	
mtu [<mtu-value>] **no mtu enable**	Enable the receiving/sending function of MTU frame. The no command disables sending and receiving function of MTU frames.

2.9 bpdu-tunnel

2.9.1 Introduction to bpdu-tunnel

BPDU Tunnel is a Layer 2 tunnel technology. It allows Layer 2 protocol packets of geographically dispersed private network users to be transparently transmitted over specific tunnels across a service provider network through modifying the destination MAC address.

2.9.1.1 bpdu-tunnel function

In MAN application, multi-branches of a corporation may connect with each other by the service provider network. VPN provided by the service provider enables the geographically dispersed networks to form a local LAN, so the service provider needs to provide the tunnel function, namely, data information generated by user's network is able to inextenso arrive at other networks of the same corporation through the service provider network. To maintain a local concept, it not only needs to transmit the data within the user's private network across the tunnel, but also transmit layer 2 protocol packets within the user's private network.

2.9.1.2 Background of bpdu-tunnel

Special lines are used in a service provider network to build user-specific Layer 2 networks. As a result, a user network is broken down into parts located at different sides of the service

provider network. As shown in Figure, User A has two devices (CE 1 and CE 2) and both devices belong to the same VLAN. User's network is divided into network 1 and network 2, which are connected by the service provider network. When Layer 2 protocol packets cannot implement the passthrough across the service provider network, the user's network cannot process independent Layer 2 protocol calculation (for example, spanning tree calculation), so they affect each other.

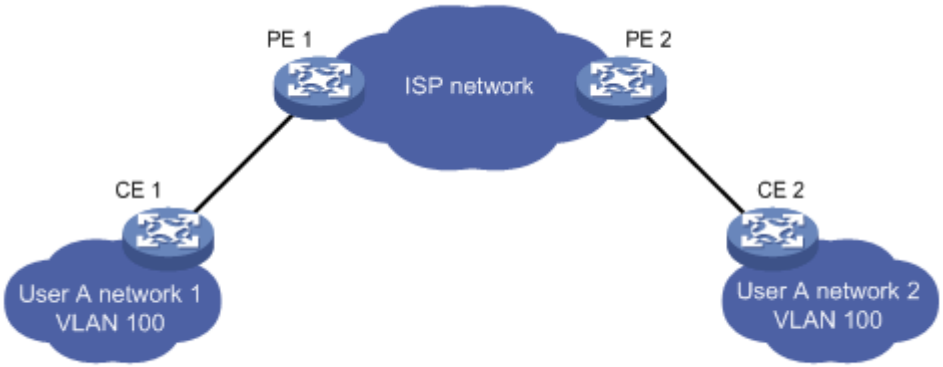


Figure 2-132 BPDU Tunnel application

2.9.2 bpdu-tunnel Configuration Task List

bpdu-tunnel configuration task list:

1. Configure tunnel MAC address globally
2. Configure the port to support the tunnel

1. Configure tunnel MAC address globally

Command	Explanation
Global mode	

bpdu-tunnel-protocol {stp| gvrp| dot1x} {group-mac <mac> | default-group-mac} **bpdu-tunnel-protocol user-defined-protocol <name> protocol-mac <mac> {group-mac <mac> | default-group-mac}** **bpdu-tunnel-protocol user-defined-protocol <name> protocol-mac <mac> encape-type ethernetii protocol-type <type> {group-mac <mac> | default-group-mac}** **bpdu-tunnel-protocol user-defined-protocol <name> protocol-mac <mac> encape-type snap {oui <oui>| } protocol-type <type> {group-mac <mac> | default-group-mac}** **bpdu-tunnel-protocol user-defined-protocol <name> protocol-mac <mac> encape-type llc dsap <dsap> ssap <ssap> {group-mac <mac> | default-group-mac}** **no bpdu-tunnel-protocol user-defined-protocol <name>**	Configure or cancel the tunnel MAC address globally.

2. Configure the port to support the tunnel

Command	Explanation
Port mode	
bpdu-tunnel-protocol {stp| gvrp| dot1x| user-defined-protocol <name>} **no bpdu-tunnel-protocol {stp| gvrp| dot1x| user-defined-protocol <name>}**	Enable the port to support the tunnel, the no command disables the function.

2.9.3 Examples of bpdu-tunnel

Special lines are used in a service provider network to build user-specific Layer 2 networks. As a result, a user network is broken down into parts located at different sides of the service provider network. As shown in Figure, User A has two devices (CE 1 and CE 2) and both devices belong to the same VLAN. User's network is divided into network 1 and network 2, which are connected by the service provider network. When Layer 2 protocol packets cannot implement the passthrough across the service provider network, the user's network cannot process independent Layer 2 protocol calculation (for example, spanning tree calculation), so they affect

each other.

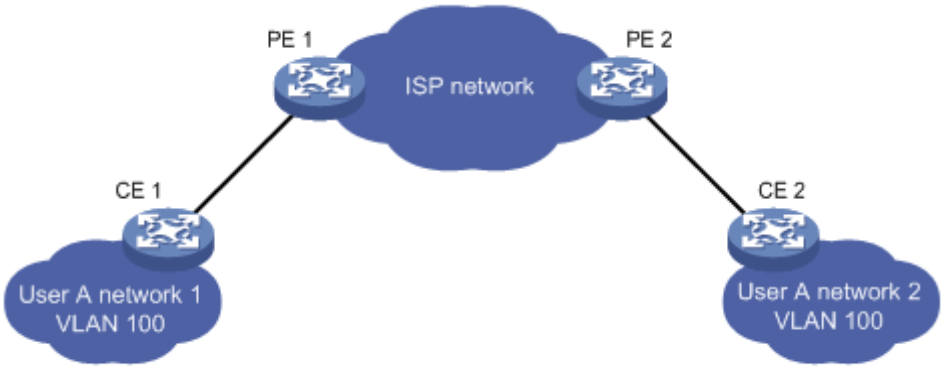


Figure 2-143 BPDU Tunnel application environment

With BPDU Tunnel, Layer 2 protocol packets from user's networks can be passed through over the service provider network in the following work flow:

1. After receiving a Layer 2 protocol packet from network 1 of user A, PE 1 in the service provider network encapsulates the packet, replaces its destination MAC address with a specific multicast MAC address, and then forwards the packet in the service provider network.
2. The encapsulated Layer 2 protocol packet (called BPDU Tunnel packet) is forwarded to PE 2 at the other end of the service provider network, which de-encapsulates the packet, restores the original destination MAC address of the packet, and then sends the packet to network 2 of user A.

The bpdu-tunnel-protocol configuration on the port of user side which is connected to the edge switches PE1 and PE2 is as below：

PE1 configuration:

PE1(config)# bpdu-tunnel-protocol stp default-group-mac

PE1(config-if-ethernet1/0/1)# bpdu-tunnel-protocol stp

PE2 configuration:

PE2(config)# bpdu-tunnel-protocol stp default-group-mac

PE2(config-if-ethernet1/0/1)# bpdu-tunnel-protocol stp

2.9.4 bpdu-tunnel Troubleshooting

After port disables stp, gvrp and dot1x functions, it is able to configure bpdu-tunnel-protocol function.

2.10 DDM

2.10.1 Introduction to DDM

2.10.1.1 Brief Introduction to DDM

DDM (Digital Diagnostic Monitor) makes the detailed digital diagnostic function standard in SFF-8472 MSA. It set that the parameter signal is monitored and make it to digitize on the circuit board of the inner module. After that, providing the demarcated result or the digitize measure result and the demarcate parameter which are saved in the standard memory framework, so as to expediently read by serial interface with double cables.

Normally, intelligent fiber modules support Digital Diagnostic function. Network management units is able to monitor the parameters (temperature, voltage, bias current, tx power and rx power) of the fiber module to obtain theirs thresholds and the real-time state of the current fiber module by the inner MCU of the fiber module. That is able to help the network management units to locate the fault in the fiber link, reduce the maintenance workload and enhance the system reliability.

DDM applications are shown in the following:

1. Module lifetime forecast

Monitoring the bias current is able to forecast the laser lifetime. Administrator is able to find some potential problems by monitoring voltage and temperature of the module.

（1）High Vcc voltage will result in the breakdown CMOS, low Vcc voltage will result in the abnormity work.

（2）High rx power will damage the receiving module, low rx power will result that the receiving module cannot work normally.

（3）High temperature will result in the fast aging of the hardware.

（4）Monitoring the received fiber power to monitor the capability of the link and the remote switch.

2. Fault location

In fiber link, locating the fault is important to the fast overload of the service, fault isolation is able to help administrator to fast locate the location of the link fault within the module (local module or remote module) or on the link, it also reduce the time for restoring the fault of the system.

Analyzing warning and alarm status of real-time parameters (temperature, voltage, bias current, tx power and rx power) can fast locate the fault through Digital Diagnostic function. Besides, the state of Tx Fault and Rx LOS is important for analyzing the fault.

3. Compatibility verification

Compatibility verification is used to analyze whether the environment of the module accords the data manual or it is compatible with the corresponding standard, because the module capability is able to be ensured only in the compatible environment. Sometimes, environment parameters exceed the data manual or the corresponding standard, it will make the falling of the

module capability that result in the transmission error.

Environment is not compatible with the module are as below:

（1）Voltage exceeds the set range

（2）Rx power is overload or is under the sensitivity of the transceiver

（3）Temperature exceeds the range of the running temperature

2.10.1.2 DDM Function

DDM descriptions are shown in the following:

1. Show the monitoring information of the transceiver

Administrator is able to know the current working state of the transceiver and find some potential problems through checking the real-time parameters (including TX power, RX power, Temperature, Voltage, Bias current) and querying the monitoring information (such as warning, alarm, real-time state and threshold, and so on). Besides, checking the fault information of the fiber module helps administrator to fast locate the link fault and saves the restored time.

2. Threshold defined by the user

For real-time parameters (TX power, RX power, Temperature, Voltage, Bias current), there are fixed thresholds. Because the user's environments are difference, the users is able to define the threshold (including high alarm, low alarm, high warn, low warn) to flexibly monitor the working state of the transceiver and find the fault directly.

The thresholds configured by the user and the manufacturer can be shown at the same time. When the threshold defined by the user is irrational, it will prompt the user and automatically process alarm or warning according to the default threshold. (the user is able to restore all thresholds to the default thresholds or restore a threshold to the default threshold)

Threshold rationality: high/low warn should be between high alarm and low alarm and high threshold should be higher than low threshold, namely, high alarm>= high warn>= low warn>= low alarm.

For fiber module, verification mode of the receiving power includes inner verification and outer verification which are decided by the manufacturer. Besides the verification mode of the real-time parameters and the default thresholds are same.

3. Transceiver monitoring

Besides checking the real-time working state of the transceiver, the user needs to monitor the detailed status, such as the former abnormity time and the abnormity type. Transceiver monitoring helps the user to find the former abnormity status through checking the log and query the last abnormity status through executing the commands. When the user finds the abnormity information of the fiber module, the fiber module information may be remonitored after processing the abnormity information, here, the user is able to know the abnormity information and renew the monitoring.

2.10.2 DDM Configuration Task List

DDM configuration task list:

1. Show the real-time monitoring information of the transceiver
2. Configure the alarm or warning thresholds of each parameter for the transceiver

3. Configure the state of the transceiver monitoring

（1） Configure the interval of the transceiver monitoring

（2） Configure the enable state of the transceiver monitoring

（3） Show the information of the transceiver monitoring

（4） Clear the information of the transceiver monitoring

1. Show the real-time monitoring information of the transceiver

Command	Explanation
User mode, admin mode and global mode	
show transceiver [interface ethernet <interface-list>][detail]	Show the monitoring of the transceiver.

2. Configure the alarm or warning thresholds of each parameter for the transceiver

Command	Explanation
Port mode	
transceiver threshold {default | {temperature | voltage | bias | rx-power | tx-power} {high-alarm | low-alarm | high-warn | low-warn} {<value> | default}}	Set the threshold defined by the user.

3. Configure the state of the transceiver monitoring

(1) Configure the interval of the transceiver monitoring

Command	Explanation
Global mode	
transceiver-monitoring interval <minutes> **no transceiver-monitoring interval**	Set the interval of the transceiver monitor. The no command sets the interval to be the default interval of 15 minutes.

（2）Configure the enable state of the transceiver monitoring

Command	Explanation
Port mode	
transceiver-monitoring {enable | disable}	Set whether the transceiver monitoring is enabled. Only the port enables the transceiver monitoring, the system records the abnormity state. After the port disables the function, the abnormity information will be clear.

（3）Show the information of the transceiver monitoring

Command	Explanation

Admin mode and global mode	
show transceiver threshold-violation [interface ethernet <interface-list>]	Show the information of the transceiver monitoring, including the last threshold-violation informatijon, the interval of the current transceiver monitoring and whether the port enables the transceiver monitoring.

（4）Clear the information of the transceiver monitoring

Command	Explanation
Admin mode	
clear transceiver threshold-violation [interface ethernet <interface-list>]	Clear the threshold violation of the transceiver monitor.

2.10.3 Examples of DDM

Example1:

Ethernet 21 and Ethernet 23 are inserted the fiber module with DDM, Ethernet 24 is inserted the fiber module without DDM, Ethernet 22 does not insert any fiber module, show the DDM information of the fiber module.

a、Show the information of all interfaces which can read the real-time parameters normally,(No fiber module is inserted or the fiber module is not supported, the information will not be shown), for example:

Switch#show transceiver

Interface	Temp（℃）	Voltage（V）	Bias（mA）	RX Power（dBM）	TX Power（dBM）
1/0/21	33	3.31	6.11	-30.54(A-)	-6.01
1/0/23	33	5.00（W+）	6.11	-20.54(W-)	-6.02

b、Show the information of the specified interface. (N/A means no fiber module is inserted or does not support the fiber module), for example:

Switch#show transceiver interface ethernet 1/0/21-22;23

Interface	Temp（℃）	Voltage（V）	Bias（mA）	RX Power（dBM）	TX Power（dBM）
1/0/21	33	3.31	6.11	-30.54(A-)	-6.01
1/0/22	N/A	N/A	N/A	N/A	N/A
1/0/23	33	5.00（W+）	6.11	-20.54(W-)	-6.02

c、Show the detailed information, including base information, parameter value of the real-time monitoring, warning, alarm, abnormity state, threshold information and the serial number, for example:

Switch#show transceiver interface ethernet 1/0/21-22;24 detail

Ethernet 1/0/21 transceiver detail information:

Base information:

SFP found in this port, manufactured by company, on Sep 29 2010.

Type is 1000BASE-SX. Serial Number is 1108000001.

Link length is 550 m for 50um Multi-Mode Fiber.
Link length is 270 m for 62.5um Multi-Mode Fiber.
Nominal bit rate is 1300 Mb/s, Laser wavelength is 850 nm.
Brief alarm information:
RX loss of signal
Voltage high
RX power low
Detail diagnostic and threshold information:

	Diagnostic	Threshold			
	Realtime Value	High Alarm	Low Alarm	High Warn	Low Warn
Temperature（℃）	33	70	0	70	0
Voltage（V）	7.31(A+)	5.00	0.00	5.00	0.00
Bias current（mA）	6.11(W+)	10.30	0.00	5.00	0.00
RX Power（dBM）	-30.54(A-)	9.00	-25.00	9.00	-25.00
TX Power（dBM）	-6.01	9.00	-25.00	9.00	-25.00

Ethernet 1/0/22 transceiver detail information: N/A

Ethernet 1/0/24 transceiver detail information:
Base information:
SFP found in this port, manufactured by company, on Sep 29 2010.
Type is 1000BASE-SX. Serial Number is 1108000001.
Link length is 550 m for 50um Multi-Mode Fiber.
Link length is 270 m for 62.5um Multi-Mode Fiber.
Nominal bit rate is 1300 Mb/s, Laser wavelength is 850 nm.
Brief alarm information: N/A
Detail diagnostic and threshold information: N/A

Explanation: If the serial number is 0, it means that it is not specified as bellow:
SFP found in this port, manufactured by company, on Sep 29 2010.
Type is 1000BASE-SX. Serial Number is not specified.
Link length is 550 m for 50um Multi-Mode Fiber.
Link length is 270 m for 62.5um Multi-Mode Fiber.
Nominal bit rate is 1300 Mb/s, Laser wavelength is 850 nm.

Example2:

Ethernet 21 is inserted the fiber module with DDM. Configure the threshold of the fiber module after showing the DDM information.

Step1: Show the detailed DDM information.
Switch#show transceiver interface ethernet 1/0/21 detail
Ethernet 1/0/21 transceiver detail information:
Base information:
……

Brief alarm information:
RX loss of signal
Voltage high
RX power low
Detail diagnostic and threshold information:

	Diagnostic			Threshold	
	Realtime Value	High Alarm	Low Alarm	High Warn	Low Warn
	---	---	---	---	---
Temperature（℃）	33	70	0	70	0
Voltage（V）	7.31(A+)	5.00	0.00	5.00	0.00
Bias current（mA）	6.11(W+)	10.30	0.00	5.00	0.00
RX Power（dBM）	-30.54(A-)	9.00	-25.00	9.00	-25.00
TX Power（dBM）	-13.01	9.00	-25.00	9.00	-25.00

Step2: Configure the tx-power threshold of the fiber module, the low-warning threshold is -12, the low-alarm threshold is -10.00.
Switch#config
Switch(config)#interface ethernet 1/0/21
Switch(config-if-ethernet1/0/21)#transceiver threshold tx-power low-warning -12
Switch(config-if-ethernet1/0/21)#transceiver threshold tx-power low-alarm -10.00

Step3: Show the detailed DDM information of the fiber module. The alarm uses the threshold configured by the user, the threshold configured by the manufacturer is labeled with the bracket. There is the alarm with 'A-' due to -13.01 is less than -12.00.
Switch#show transceiver interface ethernet 1/0/21 detail
Ethernet 1/0/21 transceiver detail information:
Base information:
……
Brief alarm information:
RX loss of signal
Voltage high
RX power low
TX power low
Detail diagnostic and threshold information:

	Diagnostic			Threshold	
	Realtime Value	High Alarm	Low Alarm	High Warn	Low Warn
	---	---	---	---	---
Temperature（℃）	33	70	0	70	0
Voltage（V）	7.31(A+)	5.00	0.00	5.00	0.00
Bias current（mA）	6.11(W+)	10.30	0.00	5.00	0.00
RX Power（dBM）	-30.54(A-)	9.00	-25.00	9.00	-25.00
TX Power（dBM）	-13.01(A-)	9.00	-12.00(-25.00)	9.00	-10.00(-25.00)

Example3:

Ethernet 21 is inserted the fiber module with DDM. Enable the transceiver monitoring of the port after showing the transceiver monitoring of the fiber module.

Step1: Show the transceiver monitoring of the fiber module. Both ethernet 21 and ethernet 22 do not enable the transceiver monitoring, its interval is set to 30 minutes.

Switch(config)#show transceiver threshold-violation interface ethernet 1/0/21-22
Ethernet 1/0/21 transceiver threshold-violation information:
Transceiver monitor is disabled. Monitor interval is set to 30 minutes.
The last threshold-violation doesn't exist.

Ethernet 1/0/22 transceiver threshold-violation information:
Transceiver monitor is disabled. Monitor interval is set to 30 minutes.
The last threshold-violation doesn't exist.

Step2: Enable the transceiver monitoring of ethernet 21.

Switch(config)#interface ethernet 1/0/21
Switch(config-if-ethernet1/0/21)#transceiver-monitoring enable

Step3: Show the transceiver monitoring of the fiber module. In the following configuration, ethernet 21 enabled the transceiver monitoring, the last threshold-violation time is Jan 02 11:00:50 2011, the detailed DDM information exceeding the threshold is also shown.

Switch(config-if-ethernet1/0/21)#quit
Switch(config)#show transceiver threshold-violation interface ethernet 1/0/21-22
Ethernet 1/0/21 transceiver threshold-violation information:
Transceiver monitor is enabled. Monitor interval is set to 30 minutes.
The current time is Jan 02 12:30:50 2011.
The last threshold-violation time is Jan 02 11:00:50 2011.
Brief alarm information:
RX loss of signal
RX power low
Detail diagnostic and threshold information:

	Diagnostic		Threshold		
	Realtime Value	High Alarm	Low Alarm	High Warn	Low Warn
	----------	----------	----------	----------	---------
Temperature（℃）	33	70	0	70	0
Voltage（V）	7.31	10.00	0.00	5.00	0.00
Bias current（mA）	3.11	10.30	0.00	5.00	0.00
RX Power（dBM）	-30.54(A-)	9.00	-25.00(-34)	9.00	-25.00
TX Power（dBM）	-1.01	9.00	-12.05	9.00	-10.00

Ethernet 1/0/22 transceiver threshold-violation information:
Transceiver monitor is disabled. Monitor interval is set to 30 minutes.
The last threshold-violation doesn't exist.

2.10.4 DDM Troubleshooting

If problems occur when configuring DDM, please check whether the problem is caused by the following reasons:

- ☞ Ensure that the transceiver of the fiber module has been inserted fast on the port, or else DDM configuration will not be shown.
- ☞ Ensure that SNMP configuration is valid, or else the warning event cannot inform the network management system.
- ☞ Because only some boards and box switches support SFP with DDM or XFP with DDM, ensure the used board and switch support the corresponding function.
- ☞ When using **show transceiver** command or **show transceiver detail** command, it cost much time due to the switch will check all ports, so it is recommended to query the monitoring information of the transceiver on the specified port.
- ☞ Ensure the threshold defined by the user is valid. When any threshold is error, the transceiver will give an alarm according to the default setting automatically.

  The DDM information of QSFP+ port does not support transmission power (TX power), it is determined by SFF-8436.
- ☞ If QSFP+ port is configured as 40G mode, it will show the DDM information of all 4 channels; if it is configured as 4x10G mode, it only shows the channel information related to the current 10G port.
- ☞ DAC wire has optical module itself, if the wire is passive mode, there is no DDM information; if the wire is active mode, DDM information can be viewed.

2.11 EFM OAM

2.11.1 Introduction to EFM OAM

Ethernet is designed for Local Area Network at the beginning, but link length and network scope is extended rapidly while Ethernet is also applied to Metropolitan Area Network and Wide Area Network along with development. Due to lack the effectively management mechanism, it affects Ethernet application to Metropolitan Area Network and Wide Area Network, implementing OAM on Ethernet becomes a necessary development trend.

There are four protocol standards about Ethernet OAM, they are 802.3ah (EFM OAM), 802.3ag (CFM), E-LMI and Y.1731. EFM OAM and CFM are set for IEEE organization. EFM OAM works in data link layer to validly discover and manage the data link status of rock-bottom. Using EFM OAM can effectively advance management and maintenance for Ethernet to ensure the stable network operation. CFM is used for monitoring the whole network connectivity and

locating the fault in access aggregation network layer. Compare with CFM, Y.1731 standard set by ITU (International Telecommunications Union) is more powerful. E-LMI standard set by MEF is only applied to UNI. So above protocols can be used to different network topology and management, between them exist the complementary relation.

EFM OAM (Ethernet in the First Mile Operation, Administration and Maintenance) works in data link layer of OSI model to implement the relative functions through OAM sublayer, figure is as bleow:

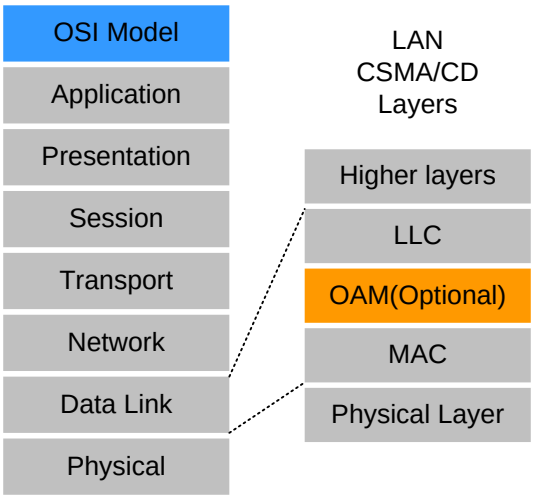


Figure 2-154 OAM location in OSI model

OAM protocol data units (OAMPDU) use destination MAC address 01-80-c2-00-00-02 of protocol, the max transmission rate is 10Pkt/s.

EFM OAM is established on the basis of OAM connection, it provides a link operation management mechanism such as link monitoring, remote fault detection and remote loopback testing, the simple introduction for EFM OAM in the following:

1. Ethernet OAM connection establishment

Ethernet OAM entity discovers remote OAM entities and establishes sessions with them by exchanging Information OAMPDUs. EFM OAM can operate in two modes: active mode and passive mode. One session can only be established by the OAM entity working in the active mode and ones working in the passive mode need to wait until it receives the connection request. After an Ethernet OAM connection is established, the Ethernet OAM entities on both sides exchange Information OAMPDUs continuously to keep the valid Ethernet OAM connection. If an Ethernet OAM entity receives no Information OAMPDU for five seconds, the Ethernet OAM connection is disconnected.

2. Link Monitoring

Fault detection in an Ethernet is difficult, especially when the physical connection in the network is not disconnected but network performance is degrading gradually. Link monitoring is used to detect and discover link faults in various environments. EFM OAM implements link monitoring through the exchange of Event Notification OAMPDUs. When detecting a link error event, the local OAM entity sends an Event Notification OAMPDU to notify the remote OAM entity. At the same time it will log information and send SNMP Trap to the network management system. While OAM entity on the other side receives the notification, it will also log and report it.

With the log information, network administrators can keep track of network status in time.

The link event monitored by EFM OAM means that the link happens the error event, including Errored symbol period event, Errored frame event, Errored frame period event, Errored frame seconds event.

Errored symbol period event: The errored symbol number can not be less than the low threshold. (Symbol: the min data transmission unit of physical medium. It is unique for coding system, the symbols may be different for different physical mediums, symbol rate means the changed time of electron status per second.)

Errored frame period event: Specifying N is frame period, the errored frame number within the period of receiving N frames can not be less than the low threshold. (Errored frame: Receiving the errored frame detected by CRC.)

Errored frame event: The number of detected error frames over M seconds can not be less than the low threshold.

Errored frame seconds event: The number of error frame seconds detected over M seconds can not be less than the low threshold. (Errored frame second: Receiving an errored frame at least in a second.)

3. Remote Fault Detection

In a network where traffic is interrupted due to device failures or unavailability, the flag field defined in Ethernet OAMPDUs allows an Ethernet OAM entity to send fault information to its peer. As Information OAMPDUs are exchanged continuously across established OAM connections, an Ethernet OAM entity can inform one of its OAM peers of link faults through Information OAMPDUs. Therefore, the network administrator can keep track of link status in time through the log information and troubleshoot in time.

There are three kinds of link faults for Information OAMPDU, they are Critical Event, Dying Gasp and Link Fault, and their definitions are different for each manufacturer, here the definitions are as below:

Critical Event: EFM OAM function of port is disabled.

Link Fault: The number of unidirectional operation or fault can not be less than the high threshold in local. Unidirectional Operation means unidirectional link can not work normally on full-duplex link without autonegotiaction. EFM OAM can detect the fault and inform the remote OAM peers through sending Information OAMPDU.

Dying Gasp: There is no definition present. Although device does not generate Dying Gasp OAMPDU, it still receives and processes such OAMPDU sent by its peer.

Typical EFM OAM application topology is in the following, it is used for point-to-point link and emulational IEEE 802.3 point-to-point link. Device enables EFM OAM through point-to-point connection to monitor the link fault in the First Mile with Ethernet access. For user, the connection between user to telecommunication is “the First Mile”, for service provider, it is “the Last Mile”.

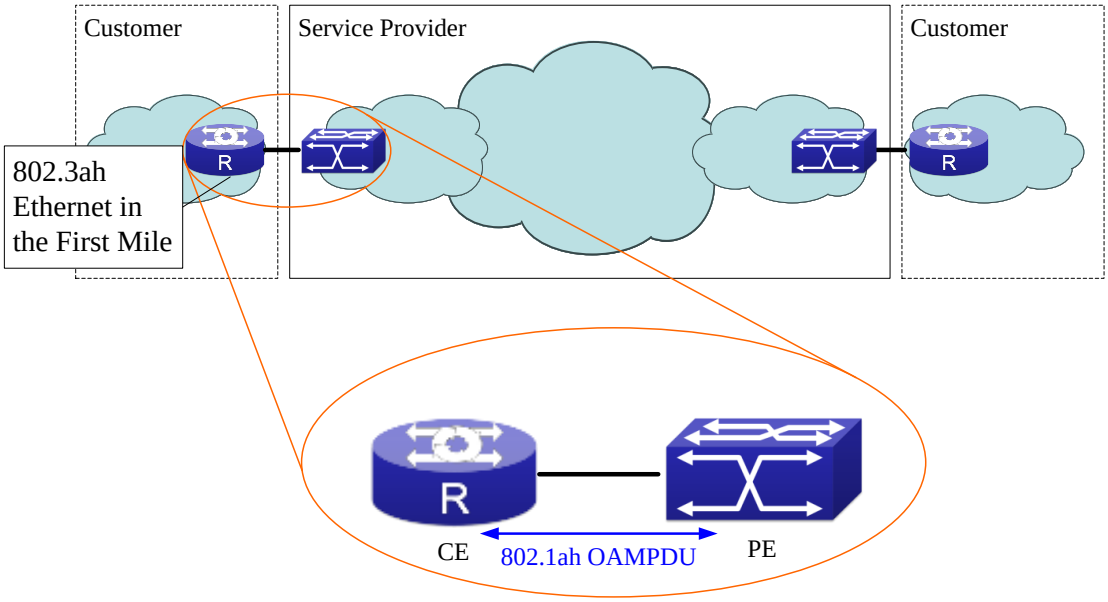


Figure 2-165 Typical OAM application topology

2.11.2 EFM OAM Configuration

EFM OAM configuration task list

1. Enable EFM OAM function of port
2. Configure link monitor
3. Configure remote failure
4. Enable EFM OAM loopback of port

Note: it needs to enable OAM first when configuring OAM parameters.

1. Enable EFM OAM function of port

Command	Explanation
Port mode	
ethernet-oam mode {active | passive}	Configure work mode of EFM OAM, default is active mode.
ethernet-oam **no ethernet-oam**	Enable EFM OAM of port, no command disables EFM OAM of port.
ethernet-oam period <seconds> **no ethernet-oam period**	Configure transmission period of OAMPDU (optional), no command restores the default value.
ethernet-oam timeout <seconds> **no ethernet-oam timeout**	Configure timeout of EFM OAM connection, no command restores the default value.

2. Configure link monitor

Command	Explanation

Port mode	
ethernet-oam link-monitor **no ethernet-oam link-monitor**	Enable link monitor of EFM OAM, no command disables link monitor.
ethernet-oam errored-symbol-period {threshold low <low-symbols> | window <seconds>} **no ethernet-oam errored-symbol-period {threshold low | window }**	Configure the low threshold and window period of errored symbol period event, no command resotores the default value. (optional)
ethernet-oam errored-frame-period {threshold low <low-frames> | window <seconds>} **no ethernet-oam errored-frame-period {threshold low | window }**	Configure the low threshold and window period of errored frame period event, no command resotores the default value.
ethernet-oam errored-frame {threshold low <low-frames> | window <seconds>} **no ethernet-oam errored-frame {threshold low | window }**	Configure the low threshold and window period of errored frame event, no command resotores the default value. (optional)
ethernet-oam errored-frame-seconds {threshold low <low-frame-seconds> | window <seconds>} **no ethernet-oam errored-frame-seconds {threshold low | window }**	Configure the low threshold and window period of errored frame seconds event, no command resotores the default value. (optional)

3. Configure remote failure

Command	Explanation
Port mode	
ethernet-oam remote-failure **no ethernet-oam remote-failure**	Enable remote failure detection of EFM OAM (failure means critical-event or link-fault event of the local), no command disables the function. (optional)
ethernet-oam errored-symbol-period threshold high {high-symbols | none} **no ethernet-oam errored-symbol-period threshold high**	Configure the high threshold of errored symbol period event, no command restores the default value. (optional)
ethernet-oam errored-frame-period threshold high {high-frames | none} **no ethernet-oam errored-frame-period threshold high**	Configure the high threshold of errored frame period event, no command restores the default value. (optional)
ethernet-oam errored-frame threshold high {high-frames | none} **no ethernet-oam errored-frame threshold high**	Configure the high threshold of errored frame event, no command restores the default value. (optional)
ethernet-oam errored-frame-seconds threshold high {high-frame-seconds | none} **no ethernet-oam errored-frame-seconds threshold high**	Configure the high threshold of errored frame seconds event, no command restores the default value. (optional)

4. Enable EFM OAM loopback of port

Command	Explanation
Port mode	
ethernet-oam remote-loopback **no ethernet-oam remote-loopback**	Enable remote EFM OAM entity to enter OAM loopback mode (its peer needs to configure OAM loopback supporting), no command cancels remote OAM loopback.

2.11.3 EFM OAM Example

Example:

CE and PE devices with point-to-point link enable EFM OAM to monitor "the First Mile" link performance. It will report the log information to network management system

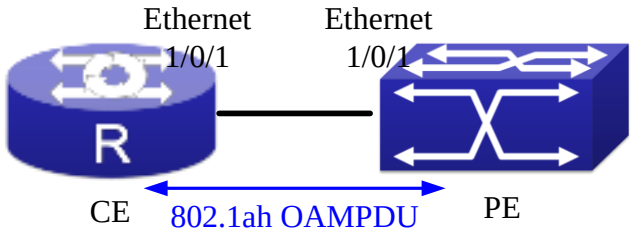


Figure 2-176 Typical OAM application topology

Configuration procedure: (Omitting SNMP and Log configuration in the following)

Configuration on CE:

CE(config)#interface ethernet1/0/1

CE (config-if-ethernet1/0/1)#ethernet-oam mode passive

CE (config-if-ethernet1/0/1)#ethernet-oam

Other parameters use the default configuration.

Configuration on PE:

PE(config)#interface ethernet 1/0/1

PE (config-if-ethernet1/0/1)#ethernet-oam

Other parameters use the default configuration.

Execute the following command when using remote loopback.

PE(config-if-ethernet1/0/1)#ethernet-oam remote-loopback

Execute the following command to make one of OAM peers exiting OAM loopback after complete detection.

PE(config-if-ethernet1/0/1)# no ethernet-oam remote-loopback

2.11.4 EFM OAM Troubleshooting

When using EFM OAM, it occurs the problem, please check whether the problem is resulted by the following reasons:

- ☞ Check whether OAM entities of two peers of link in passive mode. If so, EFM OAM connection can not be established between two OAM entities.
- ☞ Ensuring SNMP configuration is correct, or else errored event can not be reported to network management system.
- ☞ Link does not normally communicate in OAM loopback mode, it should cancel remote loopback in time after detect the link performance.
- ☞ Ensuring the used board supports remote loopback function.
- ☞ Port should not configure STP, MRPP, ULPP, Flow Control, loopback detection functions after it enables OAM loopback function, because OAM remote loopback function and these functions are mutually exclusive.

2.12 PORT SECURITY

2.12.1 Introduction to PORT SECURITY

Port security is a MAC address-based security mechanism for network access controlling. It is an extension to the existing 802.1x authentication and MAC authentication. It controls the access of unauthorized devices to the network by checking the source MAC address of the received frame and the access to unauthorized devices by checking the destination MAC address of the sent frame. With port security, you can define various port security modes to make that a device learns only legal source MAC addresses, so as to implement corresponding network security management. After port security is enabled, the device detects an illegal frame, it triggers the corresponding port security feature and takes a pre-defined action automatically. This reduces user's maintenance workload and greatly enhances system security.

2.12.2 PORT SECURITY Configuration Task List

1. Basic configuration for PORT SECURITY

Command	Explanation
Port mode	
switchport port-security **no switchport port-security**	Configure port-security of the interface.
switchport port-security mac-address <*mac-address*> [vlan <*vlan-id*>]	Configure the static security MAC of the interface.

no switchport port-security mac-address <*mac-address*> [vlan <*vlan-id*>]	
switchport port-security maximum <*value*> [vlan <*vlan-list*>] **no switchport port-security maximum <*value*> [vlan <*vlan-list*>]**	Configure the maximum number of the security MAC address allowed by the interface.
switchport port-security violation {protect | recovery | restrict | shutdown} **no switchport port-security violation**	When exceeding the maximum number of the configured MAC addresses, MAC address accessing the interface does not belongs to this interface in MAC address table or a MAC address is configured to several interfaces in same VLAN, both of them will violate the security of the MAC address.
switchport port-security aging {static | time <*value*> | type {absolute | inactivity}} **no switchport port-security violation aging {static | time | type}**	Enable port-security aging entry of the interface, specify aging time or aging type.
Admin mode	
clear port-security {all | configured | dynamic | sticky} [[address <*mac-addr*> | interface <*interface-id*>] [vlan <*vlan-id*>]]	Clear the secure MAC entry of the interface.
show port-security [interface <*interface-id*>] [address | vlan]	Show port-security configuration.

2.12.3 Example of PORT SECURITY

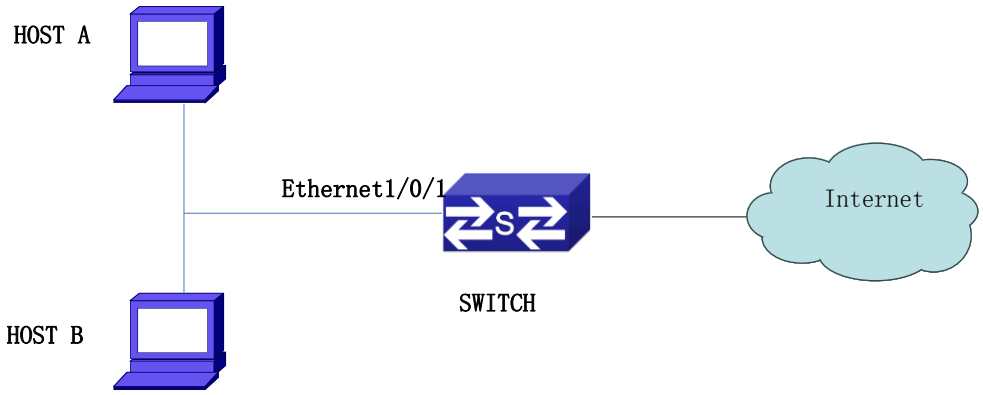


Figure 2-187 Typical topology chart for port security

When the interface enabled Port security function, configure the maximum number of the secure MAC addresses allowed by a interface to be 10, the interface allows 10 users to access the internet at most. If it exceeds the maximum number, the new user cannot access the internet, so

that it not only limit the user's number but also access the internet safely. If configuring the maximum number of the secure MAC addresses as 1, only HOST A or HOST B is able to access the internet.

Configuration process:

#Configure the switch.

Switch(config)#interface Ethernet 1/0/1

Switch(config-if-ethernet1/0/1)#switchport port-security

Switch(config-if- ethernet1/0/1)#switchport port-security maximum 10

Switch(config-if- ethernet1/0/1)#exit

Switch(config)#

2.12.4 PORT SECURITY Troubleshooting

If problems occur when configuring PORT SECURITY, please check whether the problem is caused by the following reasons:

- ☞ Check whether PORT SECURITY is enabled normally
- ☞ Check whether the valid maximum number of MAC addresses is configured

2.13 QSFP+ Port Split and Combination

2.13.1 Introduction to QSFP+ Port Split and Combination Configuration

QSFP+ port can be used as a single 40GE port and it also can be splited into 4 10GE SFP+ ports to improve the port density, decrease the user cost and increase flexibility of networks. For example, QSFP+ port of Ethernet1/1/1 is 40GE as default, it can be split into 4 10GE SFP+ ports according to need of Ethernet1/1/1, Ethernet1/1/2, Ethernet1/1/3 and Ethernet1/1/4. The SFP+ ports split support the same configuration and feature as the ordinary SFP+ physical port. QSFP+ port need the dedicated one-to-four wire to connect after split.

If user needs larger bandwidth, the 4 10GE ports can be combined to a 40GE port for using. After combination, the wire should be changed as one-to-one wire. The description of wire is shown in the relevant manual.

QSFP+ port split and combination configuration must be saved and restart the switch, then it will be effective.

2.13.2 QSFP+ Port Configuration

QSFP+ port configuration task list is as below:

1. Enter port configuration mode
2. QSFP+ port split/combination configuration
3. Configuration retained
4. Restart switch

1. QSFP+ port split/combination configuration

Command	Explanation
Global Mode	
hardware profile mode 4x10G **no hardware profile mode 4x10G**	Split the appointed QSFP+ interface from 40GE port mode into 4 SFP+ 10GE port modes; the no command combines the 4 SFP+ 10GE ports to 1 40GE port.

2.13.3 Typical Case of QSFP+ Port Configuration

Switch >en
Switch #config
Switch (config)#interface ethernet 1/1/1
Switch (config-if-ethernet1/1/1)#hardware profile mode 4x10G
The new configuration will take effect after restart!
Switch (config-if-ethernet1/1/1)#exit
Switch (config)#exit
Switch #write
Confirm to overwrite current startup-config configuration [Y/N]:y
Write running-config to current startup-config successful
Switch #%Jun 14 15:34:58 2012 Write configuration successfully!
Switch #reload

2.13.4 QSFP+ Port Configuration Troubleshooting

- Configure multiple ports split and combination successively, it will be effective only after configuring and restarting.
- For example, the name of port in 40G mode is Ethernet1/1/1, after split, it will become 4 10GE QSFP+ ports of Ethernet1/1/1, Ethernet1/1/2, Ethernet1/1/3 and Ethernet1/1/4.
- The configuration and feature of each port are the same as the ordinary physical QSFP+ port.
- After combination, use the one-to-one wire in 40G mode, both the sides are QSFP+ modules; after split, use the one-to-four wire in 4x10G mode, one side is QSFP+ mode, and the other one is divided into 4 SFP+ modules.

2.14 CFM-OAM Configuration

2.14.1 Overview

Since the Ethernet technology was naissance, its' simple and low-cost characteristics make it to become the dominant technology in the local area network.Recently, kilomega and million mege apply one after the other, this urges the network providers, facilities manfufaturers and normalizer to advance the Ethernet technology to city and wide network.

Nevertheless, Ethernet is a LAN technology constitutionally. For the need of outlying inspection, SLA (service level protocol) testing are not imminent in the LAN environment, ， therefore the trandition Ethernet does not has the OMA function which is required by the network provider. Other than that, the trandition Ethernet has 10 seconds or the shortest 1 second linkage failure rearrange time is not acceptable by the network providers. Therefore, less than 50 millisecond failure rearrange time is also a big challenge for the city network Ethernet.

After several Ethernet OMA standards (for example, IEEE 802.3ah, IEEE802.1ag and ITU Y.1731) come out, OAM is not the indication weakness of the Ethernet.Using the IEEE802.1agas example, this also go by the name of connection failure management（CFM）standard, it provides the port to port network inspection and operation tools. It can execute the tasks such as MAC Ping, L2 Trace Route etc in the huge L2 Ethernet. It will simplify the failure elimination and SLA inspection in the operation of Ethernet. At the same time, the defined CCM (continuous inspection information) in the OMA standard in Ethernet can actualize undergo the protection rearrange lower than 50 millisecond. The shortest CCM sending time in the standard is 3.3 milliseconds. Three continuous losing CCM message will cause the main link declare invalidation and using the backup link to replace. CFM's effective is built up base on the reasonable dispose of network and configuration.

2.14.1.1 Ethernet OAM Protocol Criterion

About the Ethernet OAM, there are 4 protocol standard: 802.3ah (the first mile Etherney, short form called EFM), 802.1ag (connection failure management, the short form called CFM), E-LMI (the local port of Ethernrt), Y.1731 (failure and performance inspection), constitute by different group, the corresponding relationship as following:

IEEE 802.3ah：Ethernet Link OAM (EFM OAM)

IEEE 802.1ag：Connectivity Fault Management (CFM)

ITU-Y.1731：OAM functions and mechanisms for Ethernet based network

MEF E-LMI：Ethernet Local Management Interface

EFM OAM and CFM as the constitute to set the IEEE, EFM OAMworking data link layer, as shown in Fig 14-1, can discover and manage the lower layers' data links effectively. Also, we can use EFM OAM technology to increase the management and maintance ability, in order to maintain the stability of operation. CFM is the network level of Ethernet OAM technology, mainly use for the connect pool layer and responds for inspecting the connectedness of the network, orientating the failure of network connectedness. Y.1731 is establish by the ITU, the international telecom union, its' function is mucher bigger than CFM. Can also say that the function which

perform by CFM is the subset of Y.1731. E-LMI is established by the MEF; only apply for UNI (the user boundary and the provider boundary that is faced by the user).

2.14.2 CFM OAM Basic Concept

802.1ag divide the whole network (customer, provider, operator) into different MD (Maintenance Domain). In each of the maintance domains, it wills contrapose the MD provided service for maintance management. On these services, there will have a lot of node point facilities. The core idea of the service of 802.1ag is inspecting the involved or all node points, so that it can discover the failure parts. The point that participates in maintance inspection is called MP (Maintenance Point, it is divided into MEP (maintenance edge point) and MIP (maintenance intermediate point)), the bridge of port that configure on the maintance point.

2.14.2.1 Maintance Domain

The network can be logically divided into different layers from internal to external; it is called MD (Maintenance Domain). The maintenance domain can be nesting but not across. Each of the vindicator can only see it own maintenance domain. The lower level of maintenance can provide the service to the border upon upper layer. Then, it can separate the vindicator e.g. operator and the user clearly. It can orientate the network proble more convenient.

In the network can contains of multi maintenance domain, each of the maintenance domain locates in particular level, totally there are 8 levels. The higher the level, the range of maintenance is higher. The higher level can nesting into lower level of domain. In the reaility, CFM usually applies for the following situation. At this moment, it will divide into customer domain, service provider domain, operator domain etc.

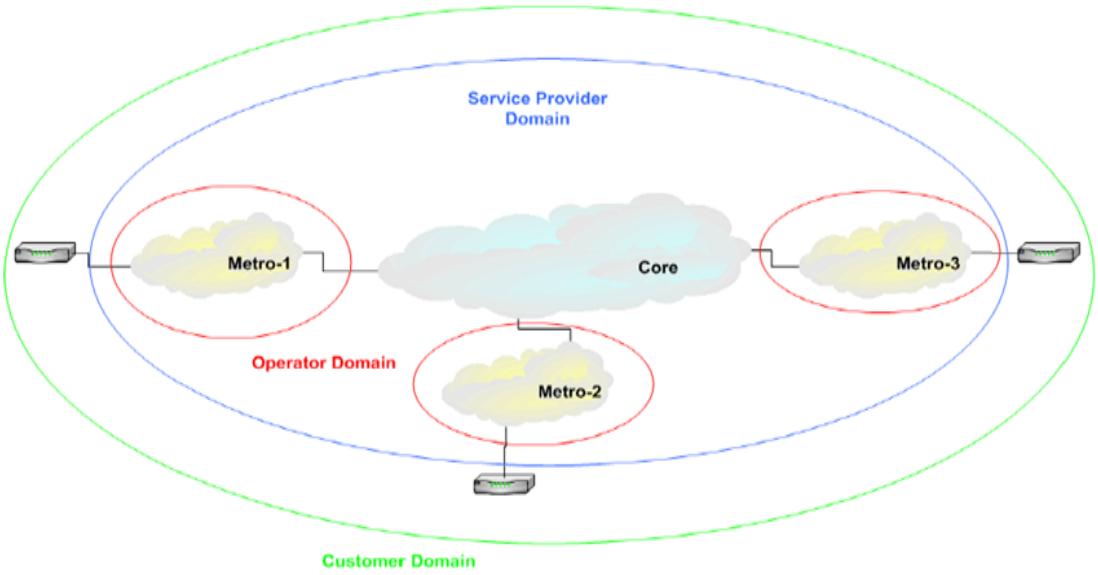


Figure 2-198 Maintenance Domain

Except the level, each of the MD has the global unique MD Name, which uses for lable that MD.

2.14.2.2 Maintenance Set

Each of the service instances in the maintenance domain is a MA (Maintenance Association). One MD usually provide several service instances externally, 802.1ag is one to one management maintenance and inspect the failure of MA. Each of the MA will have a unique name in the MD. In the network, service instance usually mark by vlanId. And MA is corresponding to the services instances. Therefore, there is a linkage between MA and vlanId. One MA corresponds to one vlanId, at the same time, protocol allows many vlanId manage and maintain by one MA. In these vlanIds, there is a vlan that is a primary vlan. In the MD, there is an uncertainty of one vlanId corresponds to multi MA situation, therefore leave out of account for this moment. All in all, md and primary vlan Id is the unique label of MA.

2.14.2.3 Maintenance Base Point

It is belong to certain maintenance service, the boundary of the service, which is configured on the port. MEP responds for initiating all CFM messages (CCM,LTM,LBM), the protocol behaviours and the status are mainly occour in the MEP. MEP is divided into UP MEP and DOWN MEP. In the bridge, if MEP is sending and receiving the MA corresponding CFM messages from Lan, then this MEP is Down MEP； if the MEP is sending and receiving the MA corresponding CFM messages from Bridge Relay, then it is the Up MEP. Shown as following Fig 14-2:

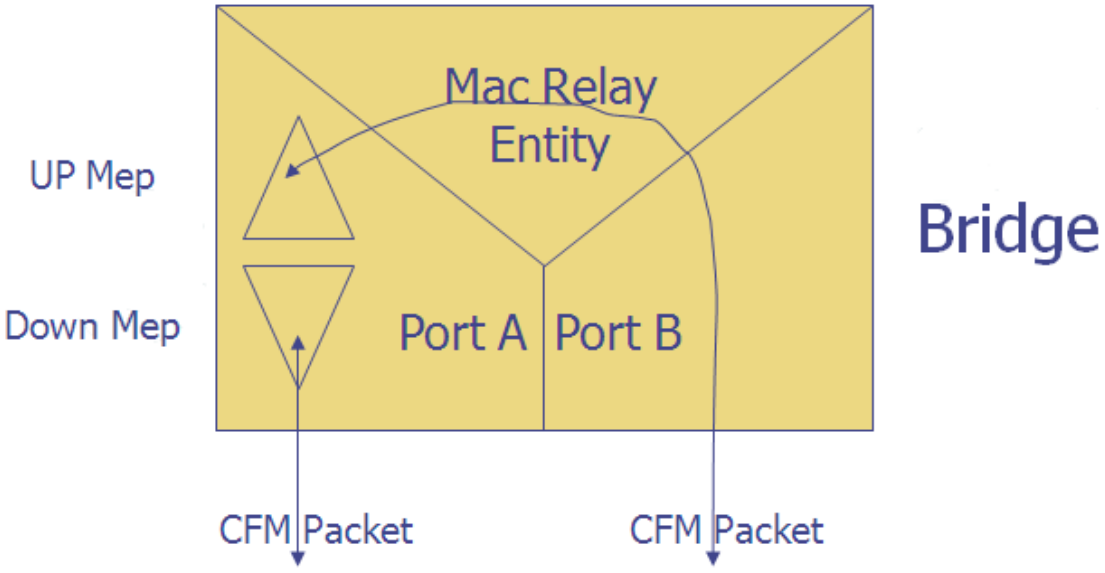


Figure 2-19 The difference between Down MEP and Up MEP

In more esay word, MEP sending and receiving the CFM messages from the local port is the Down MEP; in contrast, it is the Up MEP. MEP inherits the attributes of MD and MA that is located. It means the MD level and vlan Id. Maintenance base point is the only one label in the MA, it can be called MEPID.

2.14.2.4 Maintenance Mid-Point

It is belongs to certains maintenance service, the mid-point of maintenance service, and configures on the port. MIP cannot send the CFM messages actively; it can only receive the messages from the respond, sending Reply and transmit. It is not in charge of inspection and report failure, but it will assist MEP to undergo the failure inspection. MIP inherits the attributes of MD and MA that is located. It means the MD level and vlan Id.

MIP is not configured directly, it develops according to certain rules of system. In the port, each of the maintenance can only have one MIP. The following is the MIP rules:

none： Not build up the MIP node

default： If there is not a higher level of MEP on the port, and at the same time, lower level of MIP does not exist, then it will build up MIP on particular port at this level.

explicit： If there is not a higher level of MEP on the port, but at the same time, lower level of MEP exists and lower level of MIP does not exist, then it will build up MIP on particular port at this level.

defer： Whether build up the MIP node, the build rules will be determine by the configured rules of MD in the MA.

2.14.3 Introduction of CFM OAM Function

802.1ag provides 5 different functions: Fault detection, Fault verification and isolation, Path discovery, Fault notification and Fault recovery. Thereinto, Fault recovery need to excute with other protocols together.

2.14.3.1 Inspection of Failure

Maintanence base point （MEP）will send the CCM messages to Remote MEP in the same maintanence collection (MA) periodically. At the same time, it also receives other outlying point CCM message. If it cannot receive the CCM message in 3 months, then it will regards as occour failure in the link and report. The process is shown as follow Fig 14-3 :

CCM messages can also be sent to other MEP in the same MA, it is the group messages. The last 3 bit of the group broadcast address represent different maintanence domain level, thus it can more easy to tackle with hardware.For the lower level of MEP, after it receive higher level of messages, the hardware can accord to group broadcast MAC address and VLAN to undergo the transmitting directly.

The inspection of failure of the periodically sending CCM messages as follow:

- MEP cannot receive the CCM message on time, represent there is a failure occour in the MA;
- MEP checking the received CCM messages, it can find out the disagreement failure of the sending time interval;
- MEP received the failure MEPID or MAID, represent there is exist of internal configuration failure in MA or cross connection failure;
- MEP receive the lower level of the CCM messages, represent there is exist of internal configuration failure in MA or cross connection failure;
- MEP receive the CCM message which carries MAC status information, can investigate the outside failure of MA;

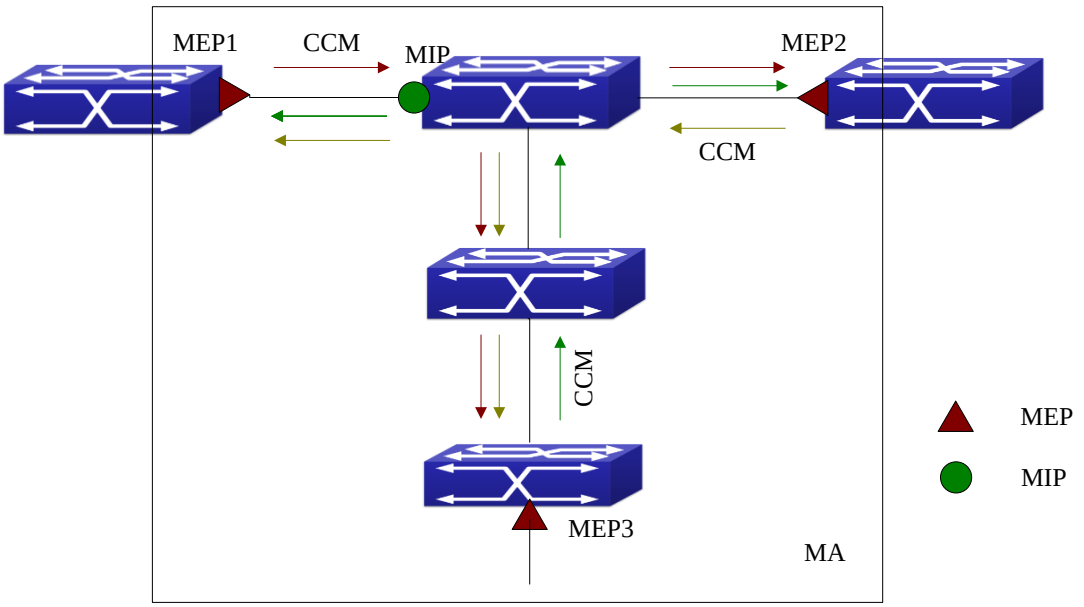


Figure 2-200 Connectedness inspection sketch map

2.14.3.2 Path Discovery

MEP and MIP will through the Linktrace to complete. The function of 802.1ag Linktrace is more or less the same with IP Traceroute. Through send the testing messages and receive replay messages to check the path of destination facilities or orientate the failure point. The processes as follow: MEP send LTM to the target MP（MEP or MIP）, each of the MIP after receiving the LTM will also send a LTR to source MEP. And then, transmit the LTM messages, until the LTM arrive to destination MP or cannot transmit at all. Source MEP according to feedback LTR to confirm the status of the linkage, and obtains the paths to target MAC. It shows in the Fig 14-4 .

LTM destination MAC address is the group broadcast address. The last 3 bit in the group broadcast address represents the level of different maintaence domain. LTR is the one way messages, the destination address as the LTM source MAC address.

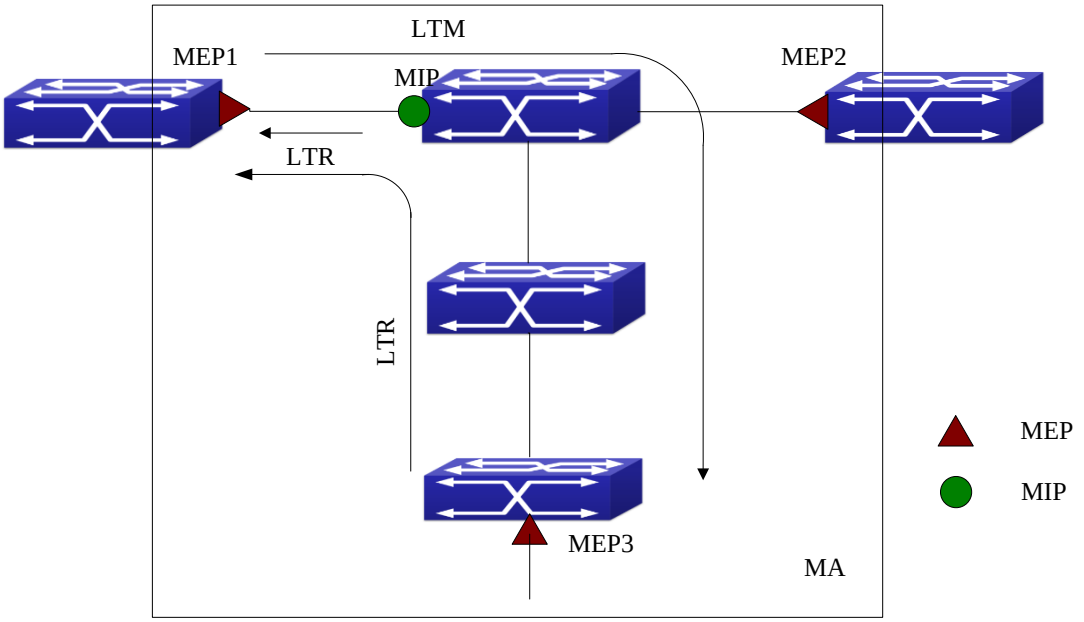


Figure 2-211 Path discovery sketch map

2.14.3.3 Confirmation and Orentation of Failure

It can be actualizing by the 802.1ag Linktrace function that mention above. Also it can be actualizing by 802.1ag Loopback function as well. The circulation function is more or less the same with IP Ping, through out the sending of testing messages and receiving the replay messages to detect whether it can arrive to the destination facility. The idiographic processes as follow: MEP sending the one way broadcast message（LBM）, the destination address of the message is the outlying MP. Once the middle facility receive the LBM will then transmit, and the ouylying MP will sending the replay message (LBR) to the source MEP after it receive the LBM. The source MEP can accord this to determine whether the outlying MP can arrive or not. As shown in the following Fig 14-5.

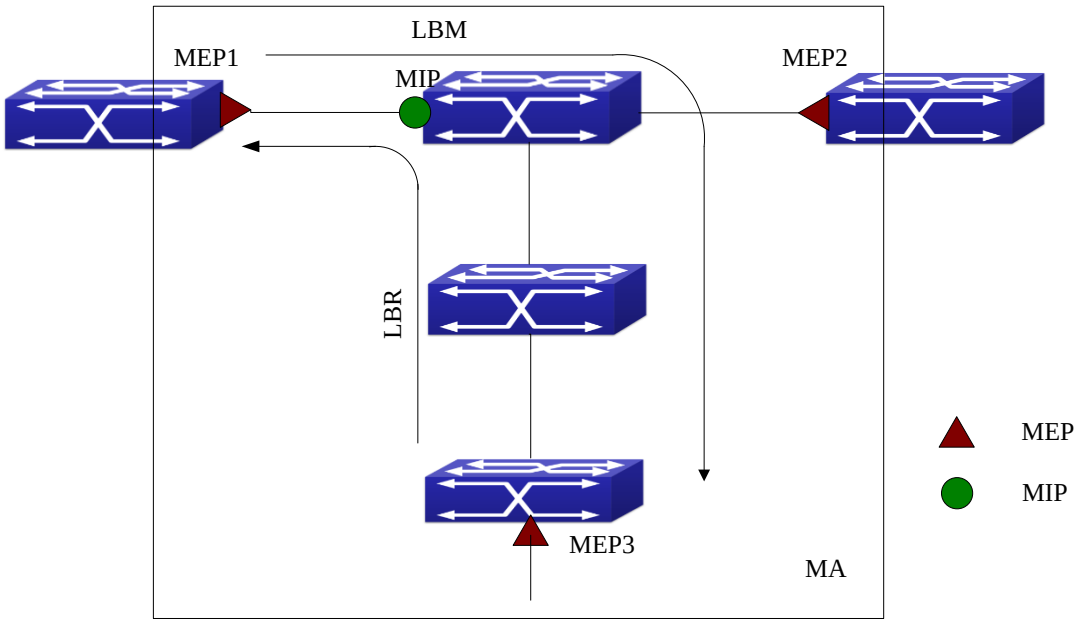


Figure 2-222 the figure of circulation function

2.14.3.4 Inform of Failure

After CFM inspects the failure of linkage, there are several of methods to tackle with:

After checking the linkage failure, MEP will send the SNMP TRAP message to the management node, and inform failure occoured;

After checking the linkage failure, MEP will send the records to the facility log book, and the administrators can discover the problem after checking it;

After checking the linkage failure, will cooperate with others automatic protect protocol such as APS etc to undergo recovery. CFM will inform the occourance of failure to these protocols, and confer switching the linkage automatically.

2.14.4 CFM OAM Basic Function Configuration

2.14.4.1 The Design of CFM Management Topology

Before the execution of CFM OAM function, need to perform the following layout in the network:

1. Need to divide the levels in the maintenance domain in the whole network, to confirm each level of boundaries in the domain.
2. Confirm the name of each maintenance domain, the name of different facilities is the same in the same maintenance domain.
3. According to the VLAN that is need to inspect, and confirm the maintenance services in different maintenance domains.
4. Confirm each of the name of all maintenance services, the name of different facilities is the same in the same maintenance sercive in the domain.
5. Confirm different facilities are the same in the same maintenance base point table in the same maintanence service of same domain.
6. Can maintance the maintenance base point under the rules of the port in the maintenance domain and boundary of services. And maintance the mid-point in the non- boundary or port.

Therefore require for the following data:

Serial number	Data
1	MD name and level
2	MA name, MA related VLAN ID
3	MEP ID, the name of the port that is connected to MEP, types of MEP
4	RMEP ID
5	MIP develop rules
6	The time interval that sending and detecting CCM news from MEP in MA

2.14.4.2 CFM OAM Configuration Task List

1. Select to enable CFM OAM function mode
2. Enable CFM OAM function globally
3. Enable y1731 function globally (selectable)
4. Create MD
5. Create MA
6. Create MEP
7. Configure RMEP
8. Create MIP (selectable)
9. Enable the failure confirmation function (selectable)
10. Configure CC sending and detecting
11. Check the configuration result of CFM

1. Select to enable CFM OAM function mode

Command	Explanation
Global Mode	
ethernet cfm mode {hw|sw|auto} **no ethernet cfm mode**	Select the mode of enabling CFM OAM; it is only used before enabling CFM OAM function. No command recovers to be the default of auto.

2. Enable CFM OAM function globally

Command	Explanation
Global Mode	
ethernet cfm global **no ethernet cfm global**	Enable CFM OAM function globally. No command disables this function.

3. Enable y1731 function globally (selectable)

Command	Explanation
Global Mode	
ethernet cfm y1731 global **no ethernet cfm y1731 global**	Open the Y1731 function. After initial this function, the switch will ente into the y1731 mode. The messages are sending and decoding in the Y1731 format. **Notice:** It need to use the ethernet cfm global command before using this command, otherwise, it cannot be function. No command disables it.

4. Create MD

Command	Explanation
Global Mode	
ethernet cfm domain < *domain-name* > level < *level-id* > ***no* ethernet cfm domain < *domain-name* >**	Build up MD，enter into the MD configuration mode. If the MD is created successfully, thelevel will not be allowed to modify. No command deletes the created MD.
MD Configuration Mode	
id {mac-address XX-XX-XX-XX-XX-XX domain-number < *domain-number* > | dns < *dns-name* > | null } **no id**	Configure MDID. Domain-name which is configured by the name of maintance domain will use the command of ***ethernet cfm domain*** will not be fill in the message. Fill in the MDID and ma name will create MAID; the total length of MAID is 44. The length cannot be existed; otherwise, it will have error. No command deletes the configured id.

5. Create MA

Command	Explanation
MD Configuration Mode	
service { < *ma-name* > | number < *ma-num* > | pvlan < *vlan-id* > }{ port | pvlan < *vlan-id* > } [vlan < *WORD* >] [direction down] **no service { < ma-name > | number < ma-num > | pvlan < vlan-id > }**	Build up MA. Configure the property of UP/DOWN of MA and enter into MA mode. One service can related to one or more vlan. If MA is created successfully, the association vlan and UP/DOWN property will not be allowed to modify. If there is need to modify, delete the MA first and create it again.One switch can configure maximum 512 MA. No command deletes the created MA.

6. Create MEP

Command	Explanation
MA Configuration Mode	
mep mepid < *WORD* > **no mep mepid [< *WORD* >]**	Using this command to build up the permit configured MEP table in the maintance collection. No command deletes the created MEP.
Port Mode	
ethernet cfm mep < *mepid* > domain < *domain-name* > service { < *ma-name* > | number < *ma-num* > | pvlan < *vlan-id* > } **no ethernet cfm mep < mepid > domain < domain-name > service { < ma-name > | number < ma-num > | pvlan < vlan-id > }**	Build up MEP on port. The MEP property has been formed when creating the MA. If MA is UP/DOWN property, all MEP points in this MA are UP-DOWN property. No command deletes the created MEP.

7. Configure RMEP

Command	Explanation
MA Configuration Mode	
continuity-check receive rmep *<mep-id>* [active time < *time* >] **no continuity-check receive rmep *<mep-id>***	Open CCM message receiving function and build up rmep in MA. If the mepid in an MA has been configured as MEP, it cannot be configured as RMEP. No command deletes the configured RMEP.

8. Create MIP (selectable)

Command	Explanation
MD Configuration Mode; MA Configuration Mode	
mip auto-create [lower-mep-only | none] **no mip auto-create**	Configure the automatic MIP in the maintance collection's domain. As default, there is no rule of configuring the mid point; and it does not carry the sender-id. No command deletes the MIP.

Global Mode	
ethernet cfm mip auto-create level < ***level-id*** **> vlan <** ***WORD*** **> [lower-mep-only] [sender-id chassis]** **no mip auto-create**	Build up the MIP configureation on the layer that does not relate to MA. As default, there is no rule of configuring the mid point; and it does not carry the sender-id. No command deletes the MIP.

9. Enable the failure confirmation function (selectable)

Command	Explanation
Global Mode	
ethernet cfm alarm {delay < ***mseconds*** **> | notification { all | error-xcon | mac-remote-error-xcon | none | remote-error-xcon | xcon} | reset <** ***mseconds >}*** **no ethernet cfm alarm { delay | notification { all | error-xcon | mac-remote-error-xcon | none | remote-error-xcon | xcon} | reset }**	Enable the function of error alarm. No command recovers to be default.
ethernet cfm logging **no ethernet cfm logging**	Open the log record function. If alarm is occour, it means that has already recorded or inform out of order. No command disables this function.
ethernet cfm snmp-server enable traps **no ethernet cfm snmp-server enable traps**	Having the snmp notification during the alarm. If the set up is success, it will have the snmp notification during the alarm. No command disables this function.

10. Configure CC sending and detecting

Command	Explanation
MA Configuration Mode	
continuity-check enable **no continuity-check enable**	Using this command to open the maintance point of CCM message sending and receiving functions. No command cancels the local CCM packets sending and detection.
continuity-check interval < ***interval-value*** **>** **no continuity-check interval**	Configure the time interval value for sending message from MEP to CCM. Under the software mode, the minimum sending cycle is 100ms (interval value=3) and under the hardware mode, the minimum sending cycle is 3.3ms (interval value=1). No command recovers to be 1s (interval value=4).

11. Check the configuration result of CFM

Command	Explanation

Admin Mode	
show ethernet cfm domain { < domain_name > | brief }	Display the configured information of maintance domain.
show ethernet cfm service [domain < *domain-name* > [service { ma-name | number < *ma-num* > | pvlan < *vlan-id* > }]]	Display the configured information of the maintance collection.
show ethernet cfm maintenance-points local [detail] [mep | mip] [domain < *domain-name* > | interface { ethernet | } <*IFNAME*>]	Display the attribute and the operation information of the maintance basepoint.
show ethernet cfm maintenance-points remote detail (mac XX-XX-XX-XX-XX-XX | domain WORD (service ((WORD)|(number <*0-65535*>)|(pvlan <*1-4094*>))) mepid <*1-4094*>)	Display the attribute and the operation information for the outlaying maintance base point.
show ethernet cfm maintenance-points remote (domain WORD (service (WORD|number <*0-65535*>|pvlan <*1-4094*>) (mepid <*1-4094*>|)|)|)	Display the attribute and operation information of outlaying maintance base point.
show ethernet cfm statistic [domain < *domain-name* > [service { ma-name | number < *ma-num* > | pvlan < *vlan-id* > }]]	Display the message sending statistics information in the CFM of the facility.

2.14.5 CFM OAM Failure Confirmation

2.14.5.1 The Confirmation of Management topology

Before excute the failure confirmation, please ensure to finish the configuration of CFM OAM function.

The following data are needed to the inspection of failure manually:

Serial number	Data
1	MD name
2	MA name
3	Destination MEP ID or MAC
4	The require number, size and overtime of sending message from the loopback function.
5	The TTL value of linktrace functionwhich need to send

2.14.5.2 Implement Loopback Function

Under the admin mode, implement the commands:

Command	Explanation
Admin Mode	
ping ethernet [target-mep < *mepid* > | target-mac < *mac-address* >] {domain < *domain-name* > service { < *ma-name* > | number < *ma-num* > | pvlan < *vlan-id* > }} [number < *number* >] [packetsize < *size* >] [timeout < *timeout* >]	Open the circulate function. Send LBM messages and receiving LBR message from a particular maintance point to the other points. Under the default stage, this function is closed. If enter into target-mep-id, it cannot searching the corresponding mac address. If it cannot find, it will display error. If you enter the mac address, then will according to this address for the circulation. If it is a domain-name, then it require opening the y1731function, then sending the group broadcast LBM message.

2.14.6 CFM OAM Failure Orientation

2.14.6.1 Management topology Confirmation

Before excute the confirmation of failure, please ensure the completion of CFM OAM function configuration.

To testing the failure manually need to prepare the following data:

Serial number	Data
1	MD name
2	MA name
3	Destation MEP ID or MAC
4 (selectable)	The largest run after reading
5 (selectable)	Only enquire for FDB; or need to enquire both FDB and MIP data for undergo the failure confirmation
6 (selectable)	MEP ID that initiates LTM

2.14.6.2 CFM OAM Failure Orientation Task List

1. Implement linktrace function

Command	Explanation
Admin Mode	
traceroute ethernet { target-mep < *target-mep-id* > | target-mac < *mac-address* > } {domain < *domain-name* > service { < *ma-name* > | number < *ma-num* > | pvlan < *vlan-id* > }} [fdb-only | source < *mepid* >]] [ttl < *ttl-value* >]	Check the path from the appointed maintaining point to the target point. As default, ttl=64 and inquiry FDB and MIP database.

2. Configure auto-linktrace function (selectable)

Command	Explanation
MA Configuration Mode	
traceroute ethernet auto **no traceroute ethernet auto**	Enable the function of sending the link track packets automatically. Enable the function of sending the link track packets automatically. As default, this function is disabled. **Notice:** After enabled this function, when the maintaining point does not receive the CCM packets from the distant point in 3.5 sending cycles of CCM packets, judge that the connection to the distant point is wrong, then send LTM packet (the target of this LTM packet is the distant maintaining point, the TTL field in LTM packet is the maximum value of 255) to locate the error through detecting the responsed LTR packet. No command disables this function.
Global Mode	
ethernet cfm auto-traceroute cache { size < ***size-value*** **> | hold-time <*minutes*>}** **no ethernet cfm auto-traceroute cache { size | hold-time }**	Configure saving the size of automatic LT detection result or over time result. As default, The buffer just records the 5 least automatic detection result, the overtime as 100minutes. No command recovers to be default.

3. View the result of auto-linktrace

Command	Explanation
Admin Mode	
show ethernet cfm traceroute-reply auto [domain < ***domain_name*** **> [service { ma-name | number <** ***ma-num*** **> | pvlan <** ***vlan-id*** **> }]]**	Display the result of the automatic LT. If there is no appointed domain, then it will display all the automatic LT result in the facilities. If there is no appointed ma, then it will display particular domain's automatic LT result in the facilities.

2.14.7 ULPP Linkage (Selectable)

2.14.7.1 ULPP Linkage Task List

1. Configure ulpp linkage

☞ Configure with the topology and ensure CC function is running normally.

- Configure the ULPP function first.
- ULPP linkage is just with down MEP.
- The vlan associated with linkage MEP must be in the protection vlan of the relevant ULPP group.

Command	Explanation
Port Mode	
switchport ulpp group *<group-id>* track cfm cc level *<level-value>*	Configure ulpp group member port to associate with cfm cc detection. When ulpp group member port received the matching cfm information (timeout or recover), conduct the association.

2. Check the result of ULPP linkage configuration

Command	Explanation
Admin Mode	
show ulpp group *<word>*	Check the result of ULPP association configuration. If the configuration is successful, the level of MD in Track-cfm-level will be shown.

2.14.7.2 Example of ULPP linkage application

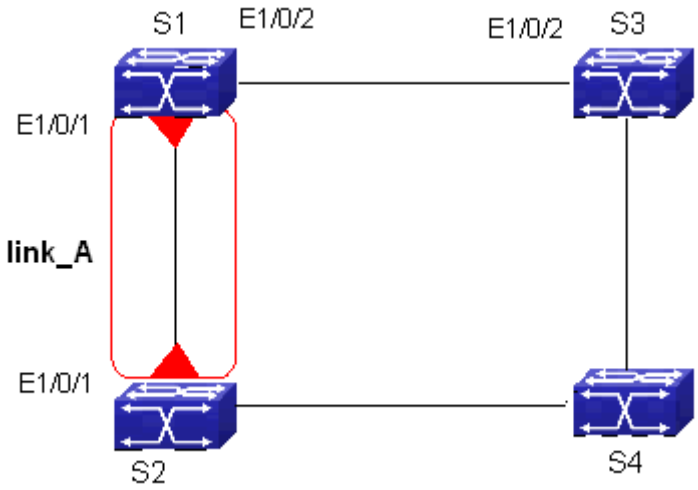


Figure 2-233 ULPP linkage application Figure

1. **The consider path of configuration:**

Using the following path of configuration to configure the ULPP linkage function:

Build up the VLAN, and adding the related ports to corresponding VLAN.

Build up the MD link_A in S1, S2 and the level is 4

Build up and configure MA1 in customer_A（MA1 and VLAN1 related）

Configure corresponding MEP and RMEP in the S1 and S2

Build up the ulpp group1 in S1 and configure the main and assist port

Configure corresponding flush receiving port to S2 and S3

Configure linkage port in the S1

2. **Steps of Configuration**

(1) Build up VLAN, and adding the related ports to corresponding VLAN

(2) Build up the MD link_A and corresponding MEP and RMEP in S1 and S2

Build up the link_A in S1 and configure corresponding MEP and RMEP

Switch(config)#ethernet cfm domain link_A level 4

Switch(config-ecfm)#service MA1 pvlan 1 direction down

Switch(config-ecfm-srv)#mep mepid 1-2

Switch(config-ecfm-srv)#continuity-check enable

Switch(config-ecfm-srv)#continuity-check receive rmep 2

Switch(config-ecfm-srv)#exit

Switch(config-ecfm)exit

Switch(config)#interface ethernet 1/0/1

Switch(config-if-ethernet1/0/1)#ethernet cfm mep 1 domain link_A service MA1

Using the same method to build up the link_A in S2 and configure corresponding MEP and RMEP

(3) Build up the ulpp group 1 in S1 and protect vlan1

Build up the ulpp group 1 in S1 and open the grabbing mode to protect vlan

Switch(config)#spanning-tree mst configuration

Switch(config-mstp-region)#instance 1 vlan 1

Switch(config-mstp-region)#exit

Switch(config)#ulpp group 1

Switch(ulpp-group-1)#protect vlan-reference-instance 1

Switch(ulpp-group-1)#flush enable mac-vlan

Switch(ulpp-group-1)#preemption mode

configure the 1/0/1 in S1 to become the master port, and the 1/0/2 as slave port

Switch(config)#interface ethernet 1/0/1

Switch(config-if-ethernet1/0/1)#ulpp group 1 master

Switch(config)#interface ethernet 1/0/2

Switch(config-if-ethernet1/0/2)#ulpp group 1 master

Configure the receiving port 1/0/1 in S2 to receive flush message

Switch(config-if-ethernet1/0/1)#ulpp flush enable mac-vlan

Using the same method to onfigure the receiving port 1/0/2 in S3 to receive flush message

(4) Configure ulpp linkage in the S1

Configure ulpp linkage in the 1/0/1 port in S1

Switch(config-if-ethernet1/0/1)#switchport ulpp group 1 track cfm level 4

3. **Checking the configuration result**

Checking the ulpp linkage configuration result in S1

Switch(config)#show ulpp group 1

ULPP group 1 information:

Description:

Preemption mode: ON

Preemption delay: 30s

Control VLAN: 1

Flush packet: MAC MAC-VLAN ARP

Protected VLAN: Reference Instance 1

Member Role State Track-cfm-level

```
--------------------------------------------------------------------------------
Ethernet1/0/1       MASTER          FORWARD             4
Ethernet1/0/2       SLAVE           STANDBY             -
--------------------------------------------------------------------------------
```

if the CFMchecking the CC is overtime, then it will inform the ULPP function to undergo the switching of main and assist ports:

%Jan 01 00:12:19 2012 ULPP: ULPP group 1:Master port Ethernet1/0/3 receives cfm event type:CFM_ALARM_RMEP_CCM vlan:1 level:4.

%Jan 01 00:12:19 2006 ULPP: ULPP group 1:Master port Ethernet1/0/3 change state to down,Slave port Ethernet1/0/1 change state to forwarding.

%Jan 01 00:12:21 2006 CFM:A CFM_ALARM_RMEP_CCM of Interface Ethernet1/0/3 is detected.

2.14.8 Example of Configuration Application

The following Fig 14-7 is the CFM configuration application illustration, in order to actualize the inspection of the status of linkage, can follow the steps as shown below to undergo the configuration.

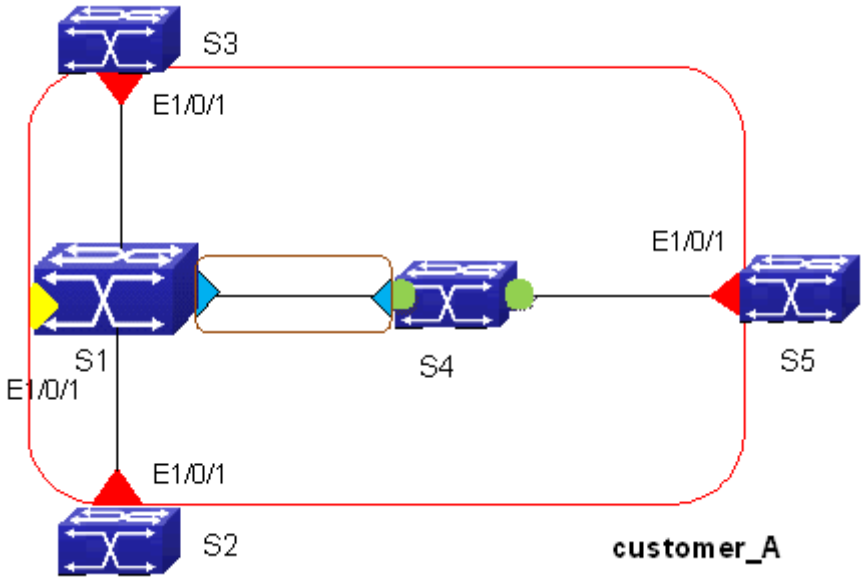


Figure 2-244 MD, MA, MEP and MIP configuration figure

1. **The consider path of configuration**

   Using the following path of configuration to configure the Ethernet CFM basic function:

   Build up the VLAN, and adding the related ports to corresponding VLAN.

   Build up the customer_A on the facilities S1、S2、S3、S4、S5 , the level of customer_A is 6.

   Build up and configure MA1 in customer_A（MA1 and VLAN1 related）

   Build up operator_A in the facilities S1, S5, the level of operator_A is 3

   Build up and configure MA2 in operator_A（MA2 and VLAN2 related）

   Build up local and outlying MEP in the MA1 of customer_A in the S1, S2, S3, and S4

   Build up MIP in the customer_A of S4

Configure local and outlying MEP in the MA2 of customer_A in the 在 S1,S5

Initial the sending and receiving function of CCM information

2. **Steps of Configuration**

(1) Build up VLAN, and adding the related ports to corresponding VLAN

(2) Open the Global CFM function and build up customer_A and configure MA1

Build up the customer_A in S1, and configure the UP direction MA1

Switch(config)#ethernet cfm global

Switch(config)#ethernet cfm domain customer_A level 6

Switch(config-ecfm)#service MA1 pvlan 1

(3) Build up the customer_A and configure the MA1 on the S2, S3, S4, and S5

Build up the customer_A in S2, and configure the Down direction MA2

Switch(config)#ethernet cfm global

Switch(config)#ethernet cfm domain customer_A level 6

Switch(config-ecfm)#service MA1 pvlan 1 direction down

Using the same method to build up the MD and MA on other facilities

(4) Build up local and outlying MEP in the MA1 of customer_A in the S1, S2, S3, and S4, also build up the MIP in S5.

#Build up MEP list as 1-4 in the MA1 of S1, configure RMEP2-4, and build up the Etherne1/0/1 on MEP1.

Switch(config-ecfm-srv)#mep mepid 1-4

Switch(config-ecfm-srv)#continuity-check receive rmep 2-4

Switch(config-ecfm-srv)exit

Switch(config-ecfm)exit

Switch(config)#interface ethernet 1/0/1

Switch(config-if-ethernet1/0/1)#ethernet cfm mep 1 domain customer_A service MA1

Build up MEP list as 1-4 in the MA1 of S2, configure RMEP1; 3-4, and build up the Etherne1/0/1 on MEP2.

Switch(config-ecfm-srv)#mep mepid 1-4

Switch(config-ecfm-srv)#continuity-check receive rmep 1； 3-4

Switch(config-ecfm-srv)exit

Switch(config-ecfm)exit

Switch(config)#interface ethernet 1/0/1

Switch(config-if-ethernet1/0/1)#ethernet cfm mep 2 domain customer_A service MA1

Build up MEP list as 1-4 in the MA1 of S3, configure RMEP1; 2; 4, and build up the Etherne1/0/1 on MEP3.

Switch(config-ecfm-srv)#mep mepid 1-4

Switch(config-ecfm-srv)#continuity-check receive rmep 1； 2； 4

Switch(config-ecfm-srv)exit

Switch(config-ecfm)exit

Switch(config)#interface ethernet 1/0/1

Switch(config-if-ethernet1/0/1)#ethernet cfm mep 3 domain customer_A service MA1

Build up MEP list as 1-4 in the MA1 of S4, configure RMEP1-3, and build up the Etherne1/0/1 on MEP3.

Switch(config-ecfm-srv)#mep mepid 1-4

Switch(config-ecfm-srv)#continuity-check receive rmep 1-3
Switch(config-ecfm-srv)exit
Switch(config-ecfm)exit
Switch(config)#interface ethernet 1/0/1
Switch(config-if-ethernet1/0/1)#ethernet cfm mep 4 domain customer_A service MA1
Using the defaultrules to build up MIP in the MA1 on S5
Switch(config-ecfm-srv)#mip auto-create

(5) Build up operator_A and configure MA2 on S1 and S5

#Build up operator_A on S1and configure MA2, the types as port
Switch(config)#ethernet cfm domain operator_A level 3
Switch(config-ecfm)#service MA2 port direction down
Using the same method (S5) mention above to build up operator_A and MA2

(6) Build up local and outlying MEP in the MA2 of operator_A in the S1 and S5

Build up MEP list as 1-2 in the S1, configure RMEP2, and build up the Etherne1/0/2 on MEP1.
Switch(config-ecfm-srv)#mep mepid 1-2
Switch(config-ecfm-srv)#continuity-check receive rmep 2
Switch(config-ecfm-srv)exit
Switch(config-ecfm)exit
Switch(config)#interface ethernet 1/0/2
Switch(config-if-ethernet1/0/1)#ethernet cfm mep 1 domain customer_A service MA2
Build up MEP list as 1-2 in the S2, configure RMEP1, and build up the Etherne1/0/2 on MEP2.
Switch(config-ecfm-srv)#mep mepid 1-2
Switch(config-ecfm-srv)#continuity-check receive rmep 1
Switch(config-ecfm-srv)exit
Switch(config-ecfm)exit
Switch(config)#interface ethernet 1/0/2
Switch(config-if-ethernet1/0/1)#ethernet cfm mep 2 domain customer_A service MA2

(7) Initial the sending and receiving function of CCM information in S1, S2, S3, S4 in MA1

Initial the sending and receiving function of CCM information in S1in MA1
Switch(config-ecfm-srv)#continuity-check receive enable
Other facilities using the same method to initial the CC function.
Initial the sending and receiving function of CCM information in S1 and S5 in MA
Switch(config-ecfm-srv)#continuity-check receive enable

(8) To check the configuration of maintanence base point of MA1 in customer_A of S1

Switch#show ethernet cfm maintenance-points local detail mep domain customer_A
Mepid:1
Port:Ethernet1/0/1 Active:1
Domain Name: customer_A
Service Name:MA1
Level:6
Vlan:1 Direction:Up

CCM:

CC Send:Enable

CC Received:Enable Interval:1(s)

2.14.9 CFM Troubleshooting

Undergo the configuration, using the CFM-OAM, it will occour errors and do not operate normally due to physical connection, configuration error.

- ☞ Ensure the whole link connection is normal, and the MA needed related vlan is existed.

1. Configuration Failure

- ☞ Ensure the system open the global OAM function, otherwise, it will failure to configure any related OAM commands
- ☞ Illegal configured MA/MD name：

MD name： 1～43 characters straing.It can be formed by letter, number, underline and the first and the last character cannot be underline.

MA name：1～43 characters straing.It can be formed by letter, number, underline and the first and the last character cannot be underline. The sum of MA and the domain name cannot excess than 44 characters.

- ☞ In the same level, a primary vlan can only be related by one MA.
- ☞ MD level, MA related pvlan and UP/DOWN attribute cannot be changed after develop
- ☞ Need to build up the MEP ID before configure the MEP and RMEP
- ☞ MEP ID in one MA , after configured as MEP, the nit cannot be change to configure RMEP
- ☞ The DOWN attribute of MA in one facility can only allow existing one MEP; one UP attribute MA can allow to exist of several of MEP, but it cannot allow to configure several identical MA MEP to the same port.

2. Cannot build up the CC connection

- ☞ Through the ***show ethernet cfm*** to checking, ensure both ports' level, MD name, MA name, MA related pvlan are the same.
- ☞ Ensure the configuration of RMEP of this port is same with the configured MEP ID at the other port.
- ☞ Ensure this port and the other port opened the CC sending and checking function
- ☞ Ensure the CCM sending period is the same for two port
- ☞ Configured the down mep on the port, then mep will receive the message from this port. If it configured the up mep, then the mep will receive the messages from others ports. Please ensure that the up mep configuration is on the non-receiving port.
- ☞ The port that is blocking by STP protocol cannot receiving, sending, replying the CFM messages. If it is configured as MEP, then even if that port was blocked by the STP protocol, it can still sending and receiving CCM messages. Only the second layer Ethernet port can support the CFM function.
- ☞ MEP and MIP can configure on the port channel, but at the same time, port channel configured MEP and MIP will ineffective on the members. If you want to increase the CFM MEP and MIP port to port channel, then that port's MEP and MIP will ineffective

as well. If you want to recover the MEP and MIP on the port, then need to delete the port-channel from this port.

3. Cannot create the MIP point

☞ Ensure the develop rules of MIP is correct

Default rules：If there is not a higher level of MEP on the port, and at the same time, lower level of MIP does not exist, then it will build up MIP on particular port at this level.

Explicit rules：If there is not a higher level of MEP on the port, but at the same time, lower level of maintanence mid-point does not exist, then it will build up the mid-point depends on whether there is a maintanence base point on lower level.

☞ It will only develop the MIP as the port status as UP; one port bases on one vlan to develop only one MIP; lower MIP point will have higher priority to develop.

☞ A DOWN attribute MA is only need to configureon the port, if there is configured the MEP point in the port, then it cannot develop the MIP, even if there is configured the port-channel, it will cause the MEP ineffective. MIP cannot develop in the MA.

If it cannot slove the proble of CFM OAM after checking, then please using the ***debug ethernet cfm*** etc command, and copy the DEBUG information (3 minutes), and then sending to the technical center of our company.

2.15 VLAN

2.15.1 Introduction to VLAN

VLAN (Virtual Local Area Network) is a technology that divides the logical addresses of devices within the network to separate network segments basing on functions, applications or management requirements. By this way, virtual workgroups can be formed regardless of the physical location of the devices. IEEE announced IEEE 802.1Q protocol to direct the standardized VLAN implementation, and the VLAN function of switch is implemented following IEEE 802.1Q.

The key idea of VLAN technology is that a large LAN can be partitioned into many separate broadcast domains dynamically to meet the demands.

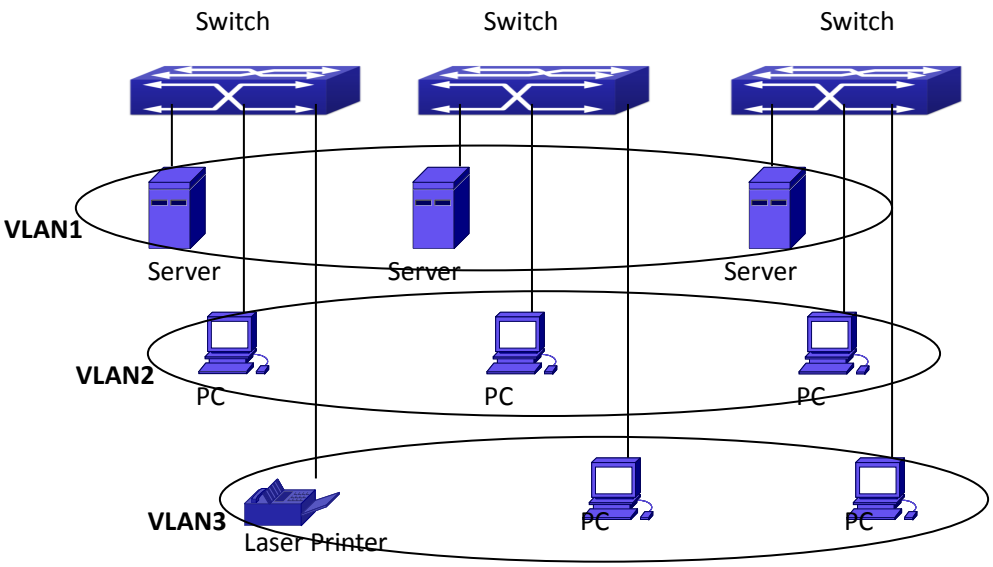


Figure 2-255 VLAN network defined logically

Each broadcast domain is a VLAN. VLANs have the same properties as the physical LANs, except VLAN is a logical partition rather than physical one. Therefore, the partition of VLANs can be performed regardless of physical locations, and the broadcast, multicast and unicast traffic within a VLAN is separated from the other VLANs.

With the aforementioned features, VLAN technology provides us with the following convenience:

- Improving network performance
- Saving network resources
- Simplifying network management
- Lowering network cost
- Enhancing network security

Switch Ethernet Ports can works in three kinds of modes: Access, Hybrid and Trunk, each mode has a different processing method in forwarding the packets with tagged or untagged.

The ports of Access type only belongs to one VLAN, usually they are used to connect the ports of the computer.

The ports of Trunk type allow multi-VLANs to pass, can receive and send the packets of multi-VLANs. Usually they are used to connect between the switches.

The ports of Hybrid type allow multi-VLANs to pass, can receive and send the packets of multi-VLANs. They can be used to connect between the switches, or to a computer of the user.

Hybrid ports and Trunk ports receive the data with the same process method, but send the data with different method: Hybrid ports can send the packets of multi-VLANs without the VLAN tag, while Trunk ports send the packets of multi-VLANs with the VLAN tag except the port native VLAN.

The switch implements VLAN and GVRP (GARP VLAN Registration Protocol) which are defined by 802.1Q. The chapter will explain the use and the configuration of VLAN and GVRP in

detail.

2.15.2 VLAN Configuration Task List

1. Create or delete VLAN
2. Set or delete VLAN name
3. Assign Switch ports for VLAN
4. Set the switch port type
5. Set Trunk port
6. Set Access port
7. Set Hybrid port
8. Enable/Disable VLAN ingress rules on ports
9. Configure Private VLAN
10. Set Private VLAN association
11. Specify internal VLAN ID

1. Create or delete VLAN

Command	Explanation
Global Mode	
vlan WORD **no vlan WORD**	Create/delete VLAN or enter VLAN Mode

2. Set or delete VLAN name

Command	Explanation
VLAN Mode	
name <vlan-name> **no name**	Set or delete VLAN name.

3. Assigning Switch ports for VLAN

Command	Explanation
VLAN Mode	
switchport interface <*interface-list*> **no switchport interface <*interface-list*>**	Assign Switch ports to VLAN.

4. Set the Switch Port Type

Command	Explanation
Port Mode	
switchport mode {trunk | access | hybrid}	Set the current port as Trunk, Access or Hybrid port.

5. Set Trunk port

Command	Explanation
Port Mode	
switchport trunk allowed vlan {WORD | all | add WORD | except WORD | remove WORD} **no switchport trunk allowed vlan**	Set/delete VLAN allowed to be crossed by Trunk. The “no” command restores the default setting.
switchport trunk native vlan <*vlan-id*> **no switchport trunk native vlan**	Set/delete PVID for Trunk port.

6. Set Access port

Command	Explanation
Port Mode	
switchport access vlan <*vlan-id*> **no switchport access vlan**	Add the current port to the specified VLAN. The “no” command restores the default setting.

7. Set Hybrid port

Command	Explanation
Port Mode	
switchport hybrid allowed vlan {WORD | all | add WORD | except WORD | remove WORD} {tag | untag} **no switchport hybrid allowed vlan**	Set/delete the VLAN which is allowed by Hybrid port with tag or untag mode.
switchport hybrid native vlan <*vlan-id*> **no switchport hybrid native vlan**	Set/delete PVID of the port.

8. Disable/Enable VLAN Ingress Rules

Command	Explanation
Port Mode	
vlan ingress enable **no vlan ingress enable**	Enable/Disable VLAN ingress rules.

9. Configure Private VLAN

Command	Explanation
VLAN mode	
private-vlan {primary | isolated | community} **no private-vlan**	Configure current VLAN to Private VLAN. The no command deletes private VLAN.

10. Set Private VLAN association

Command	Explanation
VLAN mode	

private-vlan association <secondary-vlan-list> **no private-vlan association**	Set/delete Private VLAN association.

11. Specify internal VLAN ID

Command	Explanation
Global mode	
vlan <2-4094> internal	Specify internal VLAN ID.

2.15.3 Typical VLAN Application

Scenario:

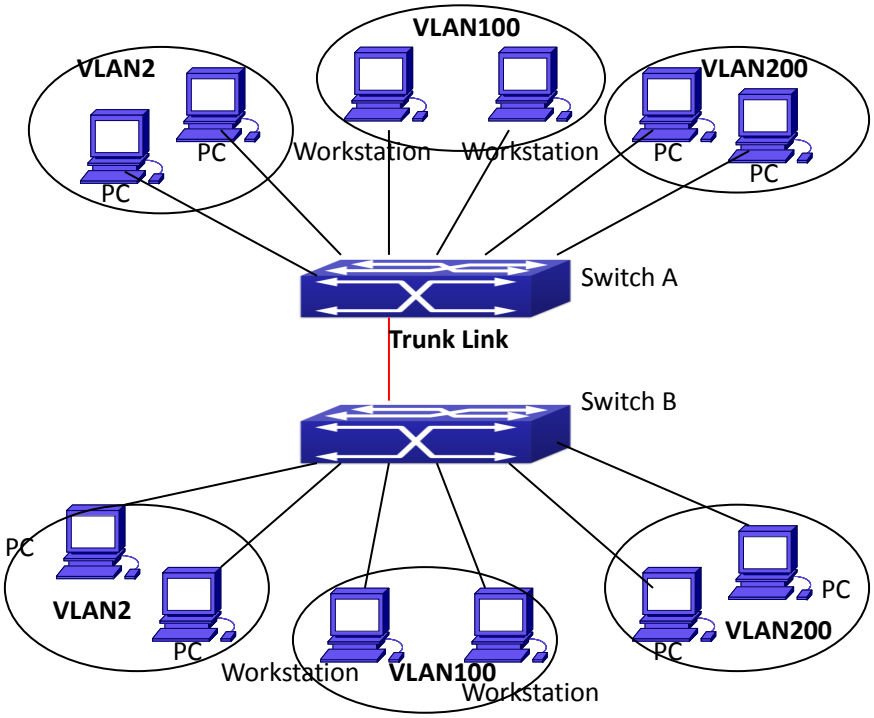


Figure 2-266 Typical VLAN Application Topology

The existing LAN is required to be partitioned to 3 VLANs due to security and application requirements. The three VLANs are VLAN2, VLAN100 and VLAN200. Those three VLANs are cross two different location A and B. One switch is placed in each site, and cross-location requirement can be met if VLAN traffic can be transferred between the two switches.

Configuration Item	Configuration description
VLAN2	Site A and site B switch port 2-4.
VLAN100	Site A and site B switch port 5-7.
VLAN200	Site A and site B switch port 8-10.
Trunk port	Site A and site B switch port 11.

Connect the Trunk ports of both switches for a Trunk link to convey the cross-switch VLAN traffic; connect all network devices to the other ports of corresponding VLANs.

In this example, port 1 and port 12 are spared and can be used for management port or for other purposes.

The configuration steps are listed below:

Switch A:

```
Switch(config)#vlan 2
Switch(Config-Vlan2)#switchport interface ethernet 1/0/2-4
Switch (Config-Vlan2)#exit
Switch (config)#vlan 100
Switch (Config-Vlan100)#switchport interface ethernet 1/0/5-7
Switch (Config-Vlan100)#exit
Switch (config)#vlan 200
Switch (Config-Vlan200)#switchport interface ethernet 1/0/8-10
Switch (Config-Vlan200)#exit
Switch (config)#interface ethernet 1/0/11
Switch (Config-If-Ethernet1/0/11)#switchport mode trunk
Switch(Config-If-Ethernet1/0/11)#exit
Switch(config)#
```

Switch B:

```
Switch(config)#vlan 2
Switch(Config-Vlan2)#switchport interface ethernet 1/0/2-4
Switch (Config-Vlan2)#exit
Switch (config)#vlan 100
Switch (Config-Vlan100)#switchport interface ethernet 1/0/5-7
Switch (Config-Vlan100)#exit
Switch (config)#vlan 200
Switch (Config-Vlan200)#switchport interface ethernet 1/0/8-10
Switch (Config-Vlan200)#exit
Switch (config)#interface ethernet 1/0/11
Switch (Config-If-Ethernet1/0/11)#switchport mode trunk
Switch (Config-If-Ethernet1/0/11)#exit
```

2.15.4 Typical Application of Hybrid Port

Scenario:

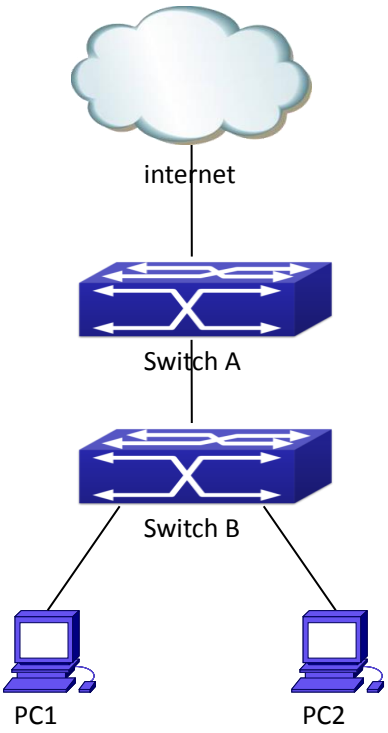


Figure 2-277 Typical Application of Hybrid Port

PC1 connects to the interface Ethernet 1/0/7 of SwitchB, PC2 connects to the interface Ethernet 1/0/9 of SwitchB, Ethernet 1/0/10 of SwitchA connect to Ethernet 1/0/10 of SwitchB.

It is required that PC1 and PC2 can not mutually access due to reason of the security, but PC1 and PC2 can access other network resources through the gateway SwitchA. We can implement this status through Hybrid port.

Configuration items are as follows:

Port	Type	PVID	the VLANs are allowed to pass
Port 1/0/10 of Switch A	Access	10	Allow the packets of VLAN 10 to pass with untag method.
Port 1/0/10 of Switch B	Hybrid	10	Allow the packets of VLAN 7, 9, 10 to pass with untag method.
Port 1/0/7 of Switch B	Hybrid	7	Allow the packets of VLAN 7, 10 to pass with untag method.
Port 1/0/9 of Switch B	Hybrid	9	Allow the packets of VLAN 9, 10 to pass with untag method.

The configuration steps are listed below:

Switch A:

Switch(config)#vlan 10

Switch(Config-Vlan10)#switchport interface ethernet 1/0/10

Switch B:

```
Switch(config)#vlan 7;9;10
Switch(config)#interface ethernet 1/0/7
Switch(Config-If-Ethernet1/0/7)#switchport mode hybrid
Switch(Config-If-Ethernet1/0/7)#switchport hybrid native vlan 7
Switch(Config-If-Ethernet1/0/7)#switchport hybrid allowed vlan 7;10 untag
Switch(Config-If-Ethernet1/0/7)#exit
Switch(Config)#interface Ethernet 1/0/9
Switch(Config-If-Ethernet1/0/9)#switchport mode hybrid
Switch(Config-If-Ethernet1/0/9)#switchport hybrid native vlan 9
Switch(Config-If-Ethernet1/0/9)#switchport hybrid allowed vlan 9;10 untag
Switch(Config-If-Ethernet1/0/9)#exit
Switch(Config)#interface Ethernet 1/0/10
Switch(Config-If-Ethernet1/0/10)#switchport mode hybrid
Switch(Config-If-Ethernet1/0/10)#switchport hybrid native vlan 10
Switch(Config-If-Ethernet1/0/10)#switchport hybrid allowed vlan 7;9;10 untag
Switch(Config-If-Ethernet1/0/10)#exit
```

2.16 GVRP

2.16.1 Introduction to GVRP

GVRP, i.e. GARP VLAN Registration Protocol, is an application of GARP (Generic Attribute Registration Protocol). GARP is mainly used to establish an attribute transmission mechanism to transmit attributes, so as to ensure protocol entities registering and deregistering the attribute. According to different transmission attributes, GARP can be divided to many application protocols, such as GMRP and GVRP. Therefore, GVRP is a protocol which transmits VLAN attributes to the whole layer 2 network through GARP protocol.

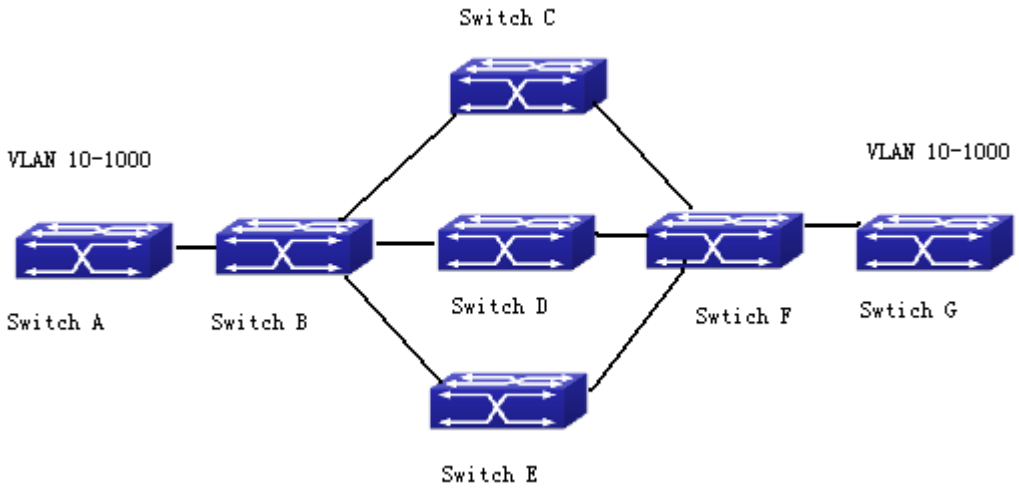


Figure 2-288 a typical application scene

A and G switches are not directly connected in layer 2 network; BCDEF are intermediate switches connecting A and G. Switch A and G configure VLAN100-1000 manually while BCDEF

switches do not. When GVRP is not enabled, A and G cannot communicate with each other, because intermediate switches without relevant VLANs. However, after GVRP is enabled on all switches, its VLAN attribute transmission mechanism enables the intermediate switches registering the VLANs dynamically, and the VLAN in VLAN100-1000 of A and G can communicate with each other. The VLANs dynamically registered by intermediate switches will be deregistered when deregistering VLAN100-1000 of A and G switches manually. So the same VLAN of two unadjacent switches can communicate mutually through GVRP protocol instead of configuring each intermediate switch manually for achieving the purpose of simplifying VLAN configuration.

2.16.2 GVRP Configuration Task List

GVRP configuration task list:
1. Configure GVRP timer
2. Configure port type
3. Enable GVRP function

1. Configure GVRP timer

Command	Explanation
Global mode	
garp timer join <200-500> **garp timer leave <500-1200>** **garp timer leaveall <5000-60000>** **no garp timer (join | leave | leaveAll)**	Configure leaveall, join and leave timer for GVRP.

2. Configure port type

Command	Explanation
Port mode	
gvrp **no gvrp**	Enable/ disable GVRP function of port.

3. Enable GVRP function

Command	Explanation
Global mode	
gvrp **no gvrp**	Enable/ disable the global GVRP function of port.

2.16.3 Example of GVRP

GVRP application:

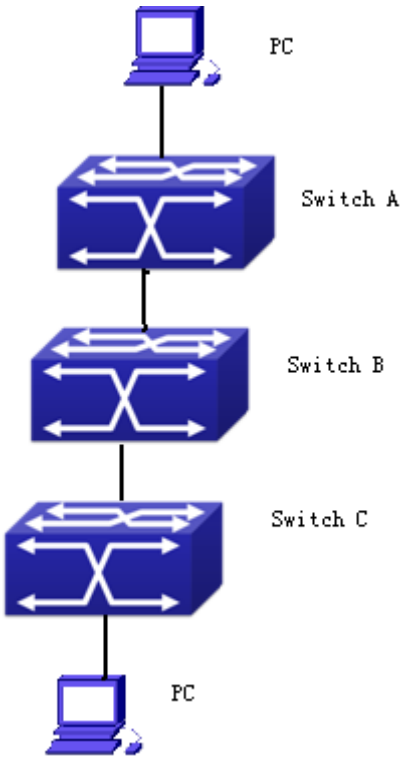


Figure 2-29 Typical GVRP Application Topology

To enable dynamic VLAN information register and update among switches, GVRP protocol is to be configured in the switch. Configure GVRP in Switch A, B and C, enable Switch B to learn VLAN100 dynamically so that two workstations connected to VLAN100 in Switch A and C can communicate with each other through Switch B without static VLAN100 entries.

Configuration Item	**Configuration description**
VLAN100	Port 2-6 of Switch A and C.
Trunk port	Port 11 of Switch A and C, Port 10, 11 of Switch B.
Global GVRP	Switch A, B, C.
Port GVRP	Port 11 of Switch A and C, Port 10, 11 of Switch B.

Connect two workstations to the VLAN100 ports in switch A and B, connect port 11 of Switch A to port 10 of Switch B, and port 11 of Switch B to port 11 of Switch C.

The configuration steps are listed below:
Switch A:
Switch(config)# gvrp
Switch(config)#vlan 100
Switch(Config-Vlan100)#switchport interface ethernet 1/0/2-6
Switch(Config-Vlan100)#exit
Switch(config)#interface ethernet 1/0/11
Switch(Config-If-Ethernet1/0/11)#switchport mode trunk
Switch(Config-If-Ethernet1/0/11)# gvrp

```
Switch(Config-If-Ethernet1/0/11)#exit

Switch B:
Switch(config)#gvrp
Switch(config)#interface ethernet 1/0/10
Switch(Config-If-Ethernet1/0/10)#switchport mode trunk
Switch(Config-If-Ethernet1/0/10)# gvrp
Switch(Config-If-Ethernet1/0/10)#exit
Switch(config)#interface ethernet 1/0/11
Switch(Config-If-Ethernet1/0/11)#switchport mode trunk
Switch(Config-If-Ethernet1/0/11)# gvrp
Switch(Config-If-Ethernet1/0/11)#exit

Switch C:
Switch(config)# gvrp
Switch(config)#vlan 100
Switch(Config-Vlan100)#switchport interface ethernet 1/0/2-6
Switch(Config-Vlan100)#exit
Switch(config)#interface ethernet 1/0/11
Switch(Config-If-Ethernet1/0/11)#switchport mode trunk
Switch(Config-If-Ethernet1/0/11)# gvrp
Switch(Config-If-Ethernet1/0/11)#exit
```

2.16.4 GVRP Troubleshooting

The GARP counter setting for Trunk ports in both ends of Trunk link must be the same, otherwise GVRP will not work normally. It is recommended to avoid enabling GVRP and RSTP at the same time in switch. If GVRP needs to be enabled, RSTP function for the ports must be disabled first.

2.17 Dot1q-tunnel

2.17.1 Introduction to Dot1q-tunnel

Dot1q-tunnel is also called QinQ (802.1Q-in-802.1Q), which is an expansion of 802.1Q. Its dominating idea is encapsulating the customer VLAN tag (CVLAN tag) to the service provider VLAN tag (SPVLAN tag). Carrying the two VLAN tags the packet is transmitted through the backbone network of the ISP internet, so to provide a simple layer-2 tunnel for the users. It is simple and easy to manage, applicable only by static configuration, and especially adaptive to small office network or small scale metropolitan area network using layer-3 switch as backbone equipment.

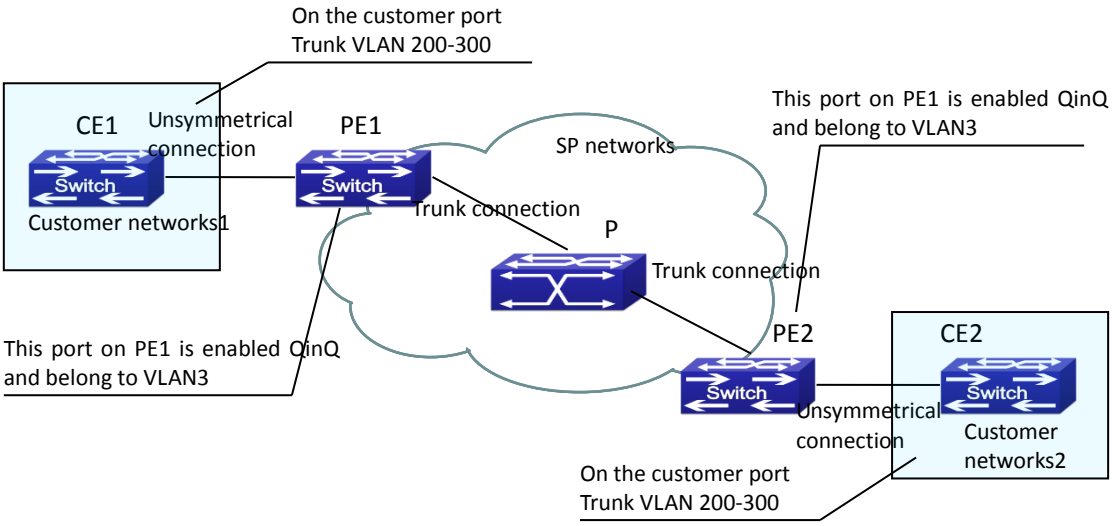


Figure 2-290 Dot1q-tunnel based Internetworking mode

As shown in above, after being enabled on the user port, dot1q-tunnel assigns each user an SPVLAN identification (SPVID). Here the identification of user is 3. Same SPVID should be assigned for the same network user on different PEs. When packet reaches PE1 from CE1, it carries the VLAN tag 200-300 of the user internal network. Since the dot1q-tunnel function is enabled, the user port on PE1 will add on the packet another VLAN tag, of which the ID is the SPVID assigned to the user. Afterwards, the packet will only be transmitted in VLAN3 when traveling in the ISP internet network while carrying two VLAN tags (the inner tag is added when entering PE1, and the outer is SPVID), whereas the VLAN information of the user network is open to the provider network. When the packet reaches PE2 and before being forwarded to CE2 from the client port on PE2, the outer VLAN tag is removed, then the packet CE2 receives is absolutely identical to the one sent by CE1. For the user, the role the operator network plays between PE1 and PE2, is to provide a reliable layer-2 link.

The technology of Dot1q-tuunel provides the ISP internet the ability of supporting many client VLANs by only one VLAN of theirselves. Both the ISP internet and the clients can configure their own VLAN independently.

It is obvious that, the dot1q-tunnel function has got following characteristics:

- ☞ Applicable through simple static configuration, no complex configuration or maintenance to be needed.
- ☞ Operators will only have to assign one SPVID for each user, which increases the number of concurrent supportable users; while the users has got the ultimate freedom in selecting and managing the VLAN IDs (select within 1~4096 at users' will).
- ☞ The user network is considerably independent. When the ISP internet is upgrading their network, the user networks do not have to change their original configuration.

Detailed description on the application and configuration of dot1q-tunnel will be provided in this section.

2.17.2 Dot1q-tunnel Configuration

Configuration Task Sequence of Dot1q-Tunnel:
1. Configure the dot1q-tunnel function on port
2. Configure the protocol type (TPID) on port

1. Configure the dot1q-tunnel function on port

Command	Explanation
Port mode	
dot1q-tunnel enable **no dot1q-tunnel enable**	Enter/exit the dot1q-tunnel mode on the port.

2. Configure the protocol type (TPID) on port

Command	Explanation
Port mode	
dot1q-tunnel tpid {0x8100|0x9100|0x9200|<1-65535>}	Configure the protocol type on TRUNK port.

2.17.3 Typical Applications of the Dot1q-tunnel

Scenario:

Edge switch PE1 and PE2 of the ISP internet forward the VLAN200~300 data between CE1 and CE2 of the client network with VLAN3. The port1 of PE1 is connected to CE1, port10 is connected to public network, the TPID of the connected equipment is 9100; port1 of PE2 is connected to CE2, port10 is connected to public network.

Configuration Item	Configuration Explanation
VLAN3	Port1 of PE1 and PE2.
dot1q-tunnel	Port1 of PE1 and PE2.
tpid	9100

Configuration procedure is as follows:

PE1:

```
Switch(config)#vlan 3
Switch(Config-Vlan3)#switchport interface ethernet 1/0/1
Switch(Config-Vlan3)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(Config-Ethernet1/0/1)# dot1q-tunnel enable
Switch(Config-Ethernet1/0/1)# exit
Switch(Config)#interface ethernet 1/0/10
Switch(Config-Ethernet1/0/10)#switchport mode trunk
Switch(Config-Ethernet1/0/10)#dot1q-tunnel tpid 0x9100
```

Switch(Config-Ethernet1/0/10)#exit
Switch(Config)#
PE2:
Switch(config)#vlan 3
Switch(Config-Vlan3)#switchport interface ethernet 1/0/1
Switch(Config-Vlan3)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(Config-Ethernet1/0/1)# dot1q-tunnel enable
Switch(Config-Ethernet1/0/1)# exit
Switch(Config)#interface ethernet 1/0/10
Switch(Config-Ethernet1/0/10)#switchport mode trunk
Switch(Config-Ethernet1/0/10)#dot1q-tunnel tpid 0x9100
Switch(Config-Ethernet1/0/10)#exit
Switch(Config)#

2.17.4 Dot1q-tunnel Troubleshooting

- Enabling dot1q-tunnel on Trunk port will make the tag of the data packet unpredictable which is not required in the application. So it is not recommended to enable dot1q-tunnel on Trunk port.

  Enabled with STP/MSTP is not supported.
- Enabled with PVLAN is not supported.

2.18 VLAN-translation

2.18.1 Introduction to VLAN-translation

VLAN translation, as one can tell from the name, which translates the original VLAN ID to new VLAN ID according to the user requirements so to exchange data across different VLANs. VLAN translation is classified to ingress translation and egress translation, respectively switch over the VLAN ID at the ingress or the egress.

Application and configuration of VLAN translation will be explained in detail in this section.

2.18.2 VLAN-translation Configuration

Configuration task sequence of VLAN-translation:
1. Configure the VLAN-translation function on the port
2. Configure the VLAN-translation relations on the port
3. Configure whether the packet is dropped when checking VLAN-translation is failing

4. Show the related configuration of vlan-translation

1. Configure the VLAN-translation of the port

Command	Explanation
Port mode	
vlan-translation enable **no vlan-translation enable**	Enter/exit the port VLAN-translation mode.

2. Configure the VLAN-translation relation of the port

Command	Explanation
Port mode	
vlan-translation <*old-vlan-id*> to <*new-vlan-id*> {in|out} **no vlan-translation old-vlan-id {in|out}**	Add/delete a VLAN-translation relation.

3. Configure whether the packet is dropped when checking VLAN-translation is failing

Command	Explanation
Port mode	
vlan-translation miss *drop* {in / out | both} ***no*** vlan-translation **miss drop {**in **| out | both}**	Configure the VLAN-translation packet dropped on port if there is any failure.

4. Show the related configuration of vlan-translation

Command	Explanation
Admin mode	
show vlan-translation	Show the related configuration of vlan-translation.

2.18.3 Typical application of VLAN-translation

Scenario:

Edge switch PE1 and PE2 of the ISP internet support the VLAN20 data task between CE1 and CE2 of the client network with VLAN3. The port1/0/1 of PE1 is connected to CE1, port1/0/10 is connected to public network; port1/0/1 of PE2 is connected to CE2, port1/0/10 is connected to public network.

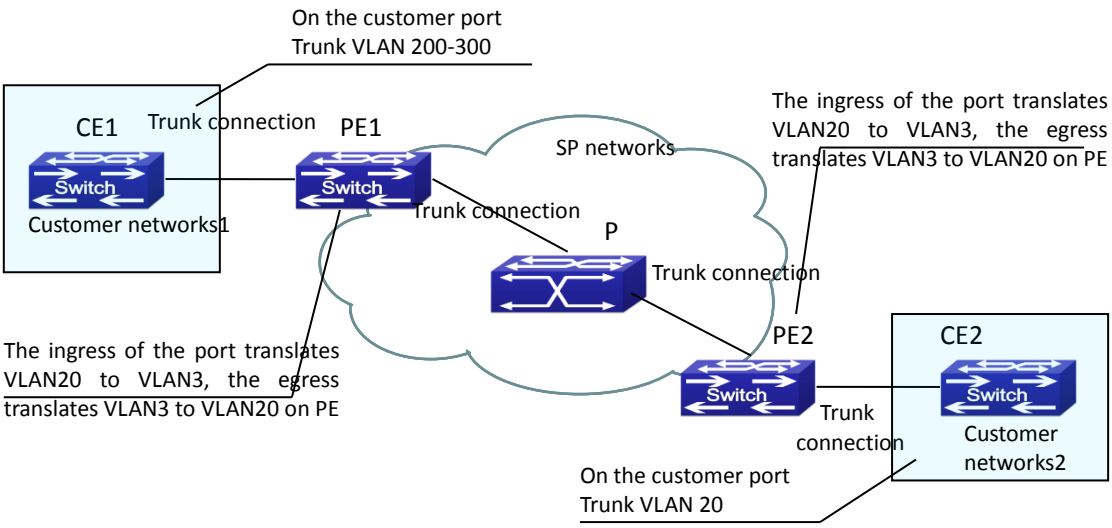


Figure 2-301 Vlan translation topology mode

Configuration Item	Configuration Explanation
VLAN-translation	Port1/0/1 of PE1 and PE2.
Trunk port	Port1/0/1 and Port1/0/10 of PE1 and PE2.

Configuration procedure is as follows:

PE1、PE2:

switch(Config)#interface ethernet 1/0/1
switch(Config-Ethernet1/0/1)#switchport mode trunk
switch(Config-Ethernet1/0/1)# vlan-translation enable
switch(Config-Ethernet1/0/1)# vlan-translation 20 to 3 in
switch(Config-Ethernet1/0/1)# vlan-translation 3 to 20 out
switch(Config-Ethernet1/0/1)# exit
switch(Config)#interface ethernet 1/0/10
switch(Config-Ethernet1/0/10)#switchport mode trunk
switch(Config-Ethernet1/0/10)#exit
switch(Config)#

2.18.4 VLAN-translation Troubleshooting

Normally the VLAN-translation is applied on trunk ports.

☞ Normally before using the VLAN-translation, the dot1q-tunnel function needs to be enabled first, to adapt double tag data packet processes VLAN-translation.

2.19 Dynamic VLAN

2.19.1 Introduction to Dynamic VLAN

The dynamic VLAN is named corresponding to the static VLAN (namely the port based VLAN). Dynamic VLAN supported by the switch includes MAC-based VLAN, IP-subnet-based VLAN and Protocol-based VLAN. Detailed description is as follows:

The MAC-based VLAN division is based on the MAC address of each host, namely every host with a MAC address will be assigned to certain VLAN. By the means, the network user will maintain his membership in his belonging VLAN when moves from a physical location to another. As we can see the greatest advantage of this VLAN division is that the VLAN does not have to be re-configured when the user physic location change, namely shift from one switch to another, which is because it is user based, not switch port based.

The IP subnet based VLAN is divided according to the source IP address and its subnet mask of every host. It assigns corresponding VLAN ID to the data packet according to the subnet segment, leading the data packet to specified VLAN. Its advantage is the same as that of the MAC-based VLAN: the user does not have to change configuration when relocated.

The VLAN is divided by the network layer protocol, assigning different protocol to different VLANs. This is very attractive to the network administrators who wish to organize the user by applications and services. Moreover the user can move freely within the network while maintaining his membership. Advantage of this method enables user to change physical position without changing their VLAN residing configuration, while the VLAN can be divided by types of protocols which is important to the network administrators. Further, this method has no need of added frame label to identify the VLAN which reduce the network traffic.

Notice: Dynamic VLAN needs to associate with Hybrid attribute of the ports to work, so the ports that may be added to a dynamic VLAN must be configured as Hybrid port.

2.19.2 Dynamic VLAN Configuration

Dynamic VLAN Configuration Task Sequence:

1. Configure the MAC-based VLAN function on the port
2. Set the VLAN to MAC VLAN
3. Configure the correspondence between the MAC address and the VLAN
4. Configure the IP-subnet-based VLAN function on the port
5. Configure the correspondence between the IP subnet and the VLAN
6. Configure the correspondence between the Protocols and the VLAN
7. Adjust the priority of the dynamic VLAN

1. Configure the MAC-based VLAN function on the port

Command	Explanation

Port Mode	
switchport mac-vlan enable **no switchport mac-vlan enable**	Enable/disable the MAC-based VLAN function on the port.

2. Set the VLAN to MAC VLAN

Command	Explanation
Global Mode	
mac-vlan vlan <*vlan-id*> **no mac-vlan**	Configure the specified VLAN to MAC VLAN; the "**no mac-vlan**" command cancels the MAC VLAN configuration of this VLAN.

3. Configure the correspondence between the MAC address and the VLAN

Command	Explanation
Global Mode	
mac-vlan mac <mac-addrss> <mac-mask> vlan <vlan-id> priority <priority-id> **no mac-vlan {mac <mac-addrss> <mac-mask>|all}**	Add/delete the correspondence between the MAC address and the VLAN, it means to make the specified MAC address join/leave the specified VLAN.

4. Configure the IP-subnet-based VLAN function on the port

Command	Explanation
Port Mode	
switchport subnet-vlan enable **no switchport subnet-vlan enable**	Enable/disable the port IP-subnet-base VLAN function on the port.

5. Configure the correspondence between the IP subnet and the VLAN

Command	Explanation
Global Mode	
subnet-vlan ip-address *<ipv4-addrss>* mask *<subnet-mask>* vlan *<vlan-id>* priority *<priority-id>* **no subnet-vlan {ip-address *<ipv4-addrss>* mask *<subnet-mask>*|all}**	Add/delete the correspondence between the IP subnet and the VLAN, namely specified IP subnet joins/leaves specified VLAN.

6. Configure the correspondence between the Protocols and the VLAN

Command	Explanation
Global Mode	
protocol-vlan mode {ethernetii etype *<etype-id>*|llc {dsap *<dsap-id>* ssap *<ssap-id>*}|snap etype *<etype-id>*} vlan	Add/delete the correspondence between the Protocols and the VLAN, namely specified protocol joins/leaves specified

<vlan-id> **priority** ***<priority-id>*** **no protocol-vlan {mode {ethernetii etype** ***<etype-id>*****|llc {dsap** ***<dsap-id>*** **ssap** ***<ssap-id>*****}|snap etype** ***<etype-id>*****}|all}**	VLAN.

7. Adjust the priority of the dynamic VLAN

Command	Explanation
Global Mode	
dynamic-vlan mac-vlan prefer **dynamic-vlan subnet-vlan prefer**	Configure the priority of the dynamic VLAN.

2.19.3 Typical Application of the Dynamic VLAN

Scenario:

In the office network Department A belongs to VLAN100. Several members of this department often have the need to move within the whole office network. It is also required to ensure the resource for other members of the department to access VLAN 100. Assume one of the members is M, the MAC address of his PC is 00-03-0f-11-22-33, when M moves to VLAN200 or VLAN300, the port connecting M is configured as Hybrid mode and belongs to VLAN100 with untag mode. In this way, the data of VLAN100 will be forwarded to the port connecting M, and implement the communication requirement in VLAN100.

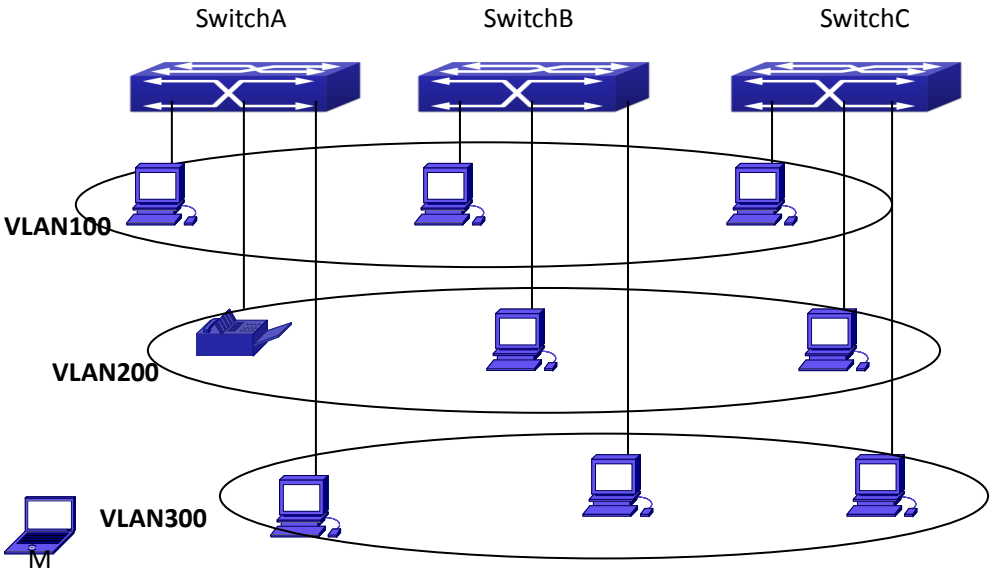


Figure 2-312 Typical topology application of dynamic VLAN

Configuration Items	**Configuration Explanation**
MAC-based VLAN	Global configuration on Switch A, Switch B, Switch C.

For example, M at E1/0/1 of SwitchA, then the configuration procedures are as follows:
Switch A, Switch B, Switch C:
SwitchA (Config)#mac-vlan mac 00-03 -0f-11-22-33 vlan 100 priority 0
SwitchA (Config)#interface ethernet 1/0/1
SwitchA (Config-Ethernet1/0/1)# swportport mode hybrid
SwitchA (Config-Ethernet1/0/1)# swportport hybrid allowed vlan 100 untagged

SwitchB (Config)#mac-vlan mac 00-03-0f-11-22-33 vlan 100 priority 0
SwitchB (Config)#exit
SwitchB#

SwitchC (Config)#mac-vlan mac 00-03-0f-11-22-33 vlan 100 priority 0
SwitchC (Config)#exit
SwitchC#

2.19.4 Dynamic VLAN Troubleshooting

- On the switch configured with dynamic VLAN, if the two connected equipment (e.g. PC) are both belongs to the same dynamic VLAN, first communication between the two equipments may not go through. The solution will be letting the two equipments positively send data packet to the switch (such as ping), to let the switch learn their source MAC, then the two equipments will be able to communicate freely within the dynamic VLAN.

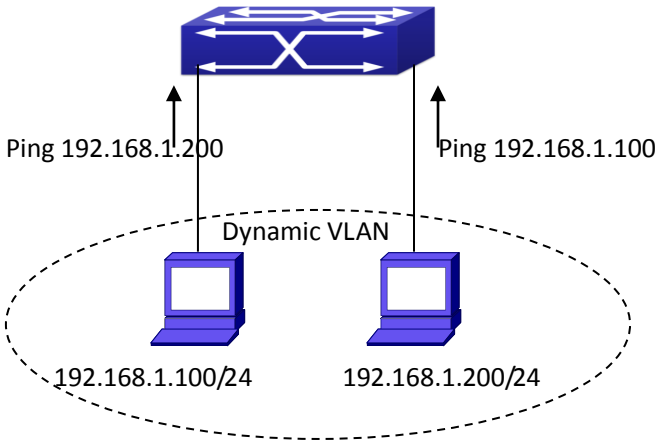


Figure 2-323 Dynamic VLAN Troubleshooting

-

2.20 Voice VLAN

2.20.1 Introduction to Voice VLAN

Voice VLAN is specially configured for the user voice data traffic. By setting a Voice VLAN and adding the ports of the connected voice equipments to the Voice VLAN, the user will be able to configure QoS (Quality of service) service for voice data, and improve the voice data traffic transmission priority to ensure the calling quality.

The switch can judge if the data traffic is the voice data traffic from specified equipment according to the source MAC address field of the data packet entering the port. The packet with the source MAC address complying with the system defined voice equipment OUI (Organizationally Unique Identifier) will be considered the voice data traffic and transmitted to the Voice VLAN.

The configuration is based on MAC address, acquiring a mechanism in which every voice equipment transmitting information through the network has got its unique MAC address. VLAN will trace the address belongs to specified MAC. By This means, VLAN allows the voice equipment always belong to Voice VLAN when relocated physically. The greatest advantage of the VLAN is the equipment can be automatically placed into Voice VLAN according to its voice traffic which will be transmitted at specified priority. Meanwhile, when voice equipment is physically relocated, it still belongs to the Voice VLAN without any further configuration modification, which is because it is based on voice equipment other than switch port.

Notice: Voice VLAN needs to associate with Hybrid attribute of the ports to work, so the ports that may be added to Voice VLAN must be configured as Hybrid port.

2.20.2 Voice VLAN Configuration

Voice VLAN Configuration Task Sequence:

1. Set the VLAN to Voice VLAN
2. Add a voice equipment to Voice VLAN
3. Enable the Voice VLAN on the port

1. Configure the VLAN to Voice VLAN

Command	Explanation
Global Mode	
voice-vlan vlan *<vlan-id>* **no voice-vlan**	Set/cancel the VLAN as a Voice VLAN

2. Add a Voice equipment to a Voice VLAN

Command	Explanation
Global Mode	
voice-vlan mac *<mac-address>* <mac-address-mask> priority <priority-id> [name *<voice-name>*] **no voice-vlan {mac *<mac-address>* *<mac-address-mask>*|name *<voice-name>***	Specify certain voice equipment join/leave the Voice VLAN

|all}	

3. Enable the Voice VLAN of the port

Command	Explanation
Port Mode	
switchport voice-vlan enable **no switchport voice-vlan enable**	Enable/disable the Voice VLAN function on the port

2.20.3 Typical Applications of the Voice VLAN

Scenario:

A company realizes voice communication through configuring Voice VLAN. IP-phone1 and IP-phone2 can be connected to any port of the switch, namely normal communication and interconnected with other switches through the uplink port. IP-phone1 MAC address is 00-03-0f-11-22-33, connect port 1/0/1 of the switch, IP-phone2 MAC address is 00-03-0f-11-22-55, connect port 1/0/2 of the switch.

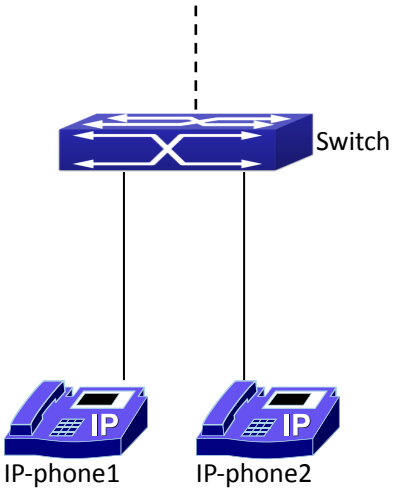


Figure 2-334 VLAN typical apply topology Figure

Configuration items	**Configuration Explanation**
Voice VLAN	Global configuration on the Switch.

Configuration procedure:

Switch 1:

Switch(config)#vlan 100

Switch(Config-Vlan100)#exit

Switch(config)#voice-vlan vlan 100

Switch(config)#voice-vlan mac 00-03-0f-11-22-33 mask 255 priority 5 name company

Switch(config)#voice-vlan mac 00-03-0f-11-22-55 mask 255 priority 5 name company

```
Switch(config)#interface ethernet 1/0/10
Switch(Config-If-Ethernet1/0/10)#switchport mode trunk
Switch(Config-If-Ethernet1/0/10)#exit
switch(Config)#interface ethernet 1/0/1
switch(Config-If-Ethernet1/0/1)#switchport mode hybrid
switch(Config-If-Ethernet1/0/1)#switchport hybrid allowed vlan 100 untag
switch(Config-If-Ethernet1/0/1)#exit
switch(Config)#interface ethernet 1/0/2
switch(Config-If-Ethernet1/0/2)#switchport mode hybrid
switch(Config-If-Ethernet1/0/2)#switchport hybrid allowed vlan 100 untag
switch(Config-If-Ethernet1/0/2)#exit
```

2.20.4 Voice VLAN Troubleshooting

Voice VLAN can not be applied concurrently with MAC-base VLAN. The Voice VLAN support maximum 1024 sets of voice equipments, the exceeded number of equipments will not be supported.

- The Voice VLAN on the port is enabled by default. If the configured data can no longer enter the Voice VLAN during operation, please check if the Voice VLAN function has been disabled on the port.

2.21 Super VLAN

2.21.1 Introduction to Super VLAN

Super VLAN, also called VLAN aggregation, we introduce the notion of super VLAN and sub VLAN. One or more sub VLANs belong to a super VLAN, and sub VLAN does not occupy an absolute subnet segment any more. However, in the same super VLAN, whether host belongs to any sub VLAN, its IP address is within the subnet segment associated with super VLAN.

In LanSwitch network, VLAN technique is applied widely for advantages (flexible control of broadcast domain and deploy convenience). However, in generic layer-3 switch, it implements communication between broadcast domains through a method that a VLAN correspond to a layer-3 interface, it results in IP address waste.

For example, VLAN division of device is shown in the figure.

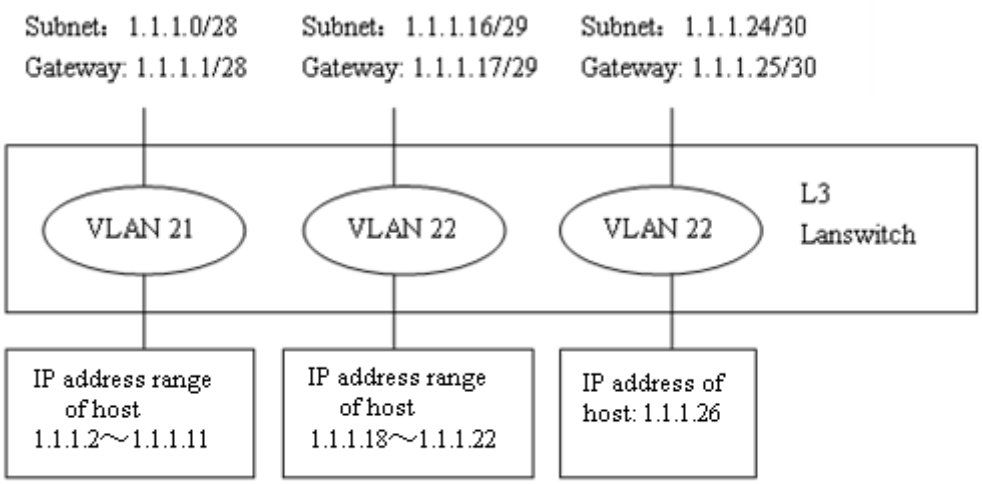


Figure 2-35 generic VLAN network

VLAN	IP Subnet	Gateway Address	Usable Hosts	Customer Hosts	Needed Hosts
21	1.1.1.0/28	1.1.1.1	14	13	10
22	1.1.1.16/29	1.1.1.17	6	5	5
23	1.1.1.24/30	1.1.1.25	2	1	1

Tab 1-1 host address division of generic VLAN

In above table, the needed hosts may be 10 in VLAN 21 and assign a subnet with mask of 28 bits—1.1.1.0/28. However, subnet 1.1.1.0 of network segment, subnet broadcast address 1.1.1.15 and the default gateway address 1.1.1.1 can not become the host address, address range within 1.1.1.2 to 1.1.1.14 can become the host address. So the usable host addresses total is 13 (2^{32-28}-3=13), however, only 10 addresses can satisfy the requirement for VLAN21.

The rest may be deduced by analogy, the needed host addresses is 5 for VLAN 22. To satisfy request, it needs to assign a subnet with mask of 29 bits (1.1.1.16/29) for VLAN 22. The needed host address is only 1 for VLAN 23, but it occupies subnet 1.1.1.24/30.

The needed addresses is 16 (10+5+1) for three VLANs, but the excellent method also occupies 28 addresses ($2^{32-28}+2^{32-29}+2^{32-30}$=28) according to address division of generic VLAN, so waste many addresses. Besides, if the needed hosts only reach to 3, but not to 10 hosts, and a subnet with mask of 28 bits is assigned to VLAN 21 before, here, the redundant addresses are wasted because they can not be used by other VLAN.

The above division is difficult to network update. If client of VLAN23 needs to add 2 hosts and client does not want to change the assigned IP address, but the addresses after 1.1.1.24 are assigned to others, so we should assign a subnet with mask of 29 bits and a new VLAN to this client. Here, there are 3 hosts for client of VLAN23 only, but this client is assigned two subnets (they are not in the same VLAN), therefore, it is difficult to management.

We can see that the number of the wasting IP address (such as subnet number, broadcast address, default gateway address) is considerable and badly reduce the addressing flexility that waste many addresses. Therefore, Super VLAN is developed for solving the problem.

Super VLAN advantages are shown in the following:
Reduce IP address number about subnet number, default gateway address and broadcast address
Implement that the different broadcast domains use addresses of the same subnet segment
Enhance addressing flexility
Reduce the address waste

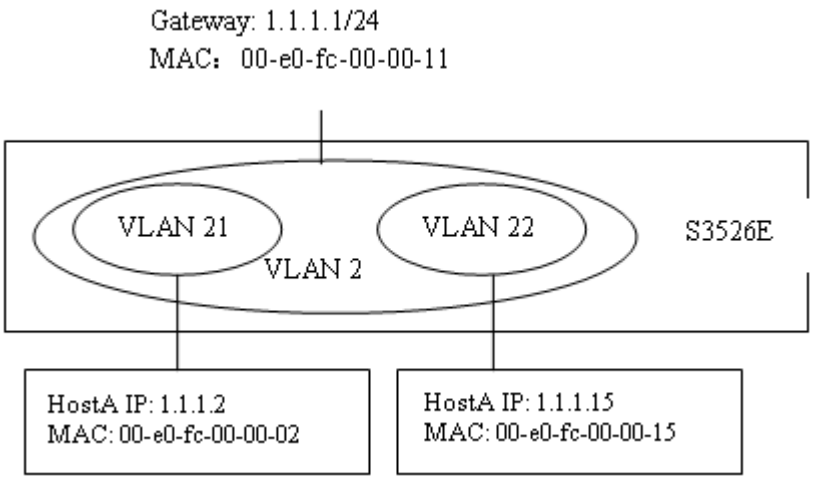


Figure 2-36 super vlan function

Super VLAN is different to the generic VLAN. Super VLAN only create a layer-3 interface and does not include any ports, is a layer-3 notion.
Layer-3 interface of super VLAN is also at UP state as long as there is physical port is at UP state in its sub-VLAN.

2.21.2 Super VLAN Configuration

1. Create or delete supervlan
2. Specify or delete subvlan
3. Enable or disable arp-proxy function of subvlan
4. Specify or delete ip-addr-range of interface
5. Specify or delete ip-addr-range of subvlan

1. Create or delete supervlan

Command	Explanation
VLAN configuration mode	
supervlan **no supervlan**	Create or delete supervlan.

2. Specify or delete subvlan

Command	Explanation
VLAN configuration mode	
subvlan WORD **no subvlan {WORD | all}**	Specify or delete subvlan.

3. Enable or disable arp-proxy function of subvlan

Command	Explanation
Interface configuration mode	
arp-proxy subvlan {WORD | all} **no arp-proxy subvlan {WORD | all}**	Enable or disable arp-proxy function of subvlan.

4. Specify or delete ip-addr-range of interface

Command	Explanation
Interface configuration mode	
ip-addr-range <*ipv4-addrss*> to <*ipv4-addrss*> **no ip-addr-range**	Specify or delete address range of interface.

5. Specify or delete ip-addr-range of subvlan

Command	Explanation
Interface configuration mode	
ip-addr-range subvlan <*vlan-id*> <*ipv4-addrss*> to <*ipv4-addrss*> **no ip-addr-range subvlan <*vlan-id*>**	Specify or delete address range of subvlan.

2.21.3 Typical Application of Super VLAN

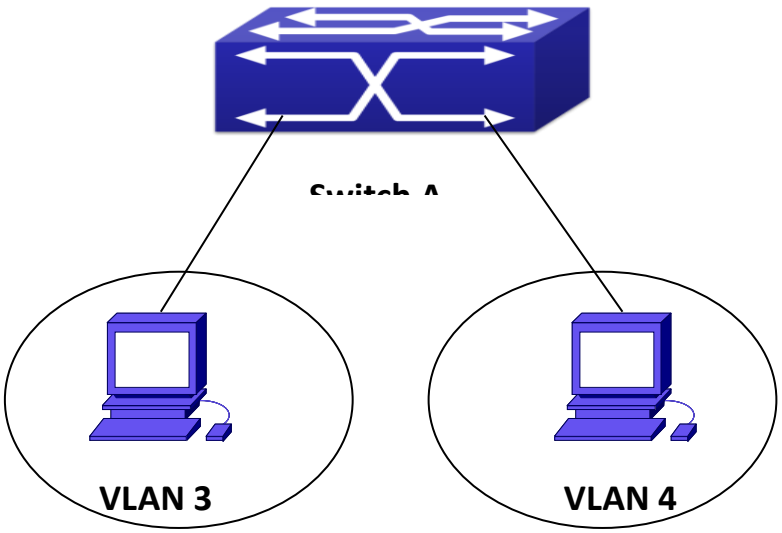


Figure 2-347 typical of super vlan topology

Terminals of two VLANs needs to configure their addresses in the same network segment due to requirement of LAN application. Layer-2 flows needs to be isolated, but layer-3 flows can be forwarded normally between two terminals. Address range of VLAN3 from 1.1.1.1 to 1.1.1.10, address range of VLAN4 from 1.1.1.20 to 1.1.1.30, layer-3 flows of terminals within two address ranges allows to be forwarded only. To implement this requirement, it needs to configure

supervlan on switch.

Configuration items	Configuration Explanation
VLAN2	Supervlan
VLAN3	Port1 of switchA
VLAN4	Port2 of switchA

Configuration procedure:

Switch A:

```
switch(Config)#vlan 2-4
switch(Config)#vlan 2
switch(Config-Vlan2)#supervlan
switch(Config-Vlan2)#subvlan 3;4
switch(Config-Vlan2)#exit
switch(Config)#interface vlan 2
switch(config-if-vlan2)#ip address 1.1.1.254 255.255.255.0
switch(config-if-vlan2)#arp-proxy subvlan all
switch(config-if-vlan2)#ip-addr-range subvlan 3 1.1.1.1 to 1.1.1.10
switch(config-if-vlan2)#ip-addr-range subvlan 4 1.1.1.20 to 1.1.1.30
switch(config-if-vlan2)#exit
```

2.21.4 Super VLAN Troubleshooting

- Supervlan and these functions (VRRP, dynamic VLAN, private VLAN, multicast VLAN, etc.) are mutually exclusive, so they should not be used at the same time.
- arp-proxy of subvlan takes effect to only one subvlan. Flow received by VLAN with arp-proxy can be forwarded to other VLAN. When two devices send flows to each other at different subvlan, please enable arp-proxy function on two subvlans.
- Sub vlan can not set layer-3 interface.
- When creating/deleting supervlan, VLAN needs to ensure no layer-3 interface, if not, it will result in error.
- If interface of supervlan specifies IP address range, but do not specify address range of subvlan, address range set by interface is standard. If both interface and subvlan specify IP address range, please check whether packets are within subvlan address range firstly, second, check whether packets are within interface address range. After that, packets can be processed other operations.
- When setting supervlan or subvlan, VLAN must be existed, it can be set.
- When port mode is set as trunk, it will automatically filter supervlan from allow-vlan.

2.22 MAC Address Table

2.22.1 Introduction to MAC Table

MAC table is a table identifies the mapping relationship between destination MAC addresses and switch ports. MAC addresses can be categorized as static MAC addresses and dynamic MAC addresses. Static MAC addresses are manually configured by the user, have the highest priority and are permanently effective (will not be overwritten by dynamic MAC addresses); dynamic MAC addresses are entries learnt by the switch in data frame forwarding, and is effective for a limited period. When the switch receives a data frame to be forwarded, it stores the source MAC address of the data frame and creates a mapping to the destination port. Then the MAC table is queried for the destination MAC address, if hit, the data frame is forwarded in the associated port, otherwise, the switch forwards the data frame to its broadcast domain. If a dynamic MAC address is not learnt from the data frames to be forwarded for a long time, the entry will be deleted from the switch MAC table.

There are two MAC table operations:

1. Obtain a MAC address.
2. Forward or filter data frame according to the MAC table.

2.22.1.1 Obtaining MAC Table

The MAC table can be built up statically and dynamically. Static configuration is to set up a mapping between the MAC addresses and the ports; dynamic learning is the process in which the switch learns the mapping between MAC addresses and ports, and updates the MAC table regularly. In this section, we will focus on the dynamic learning process of MAC table.

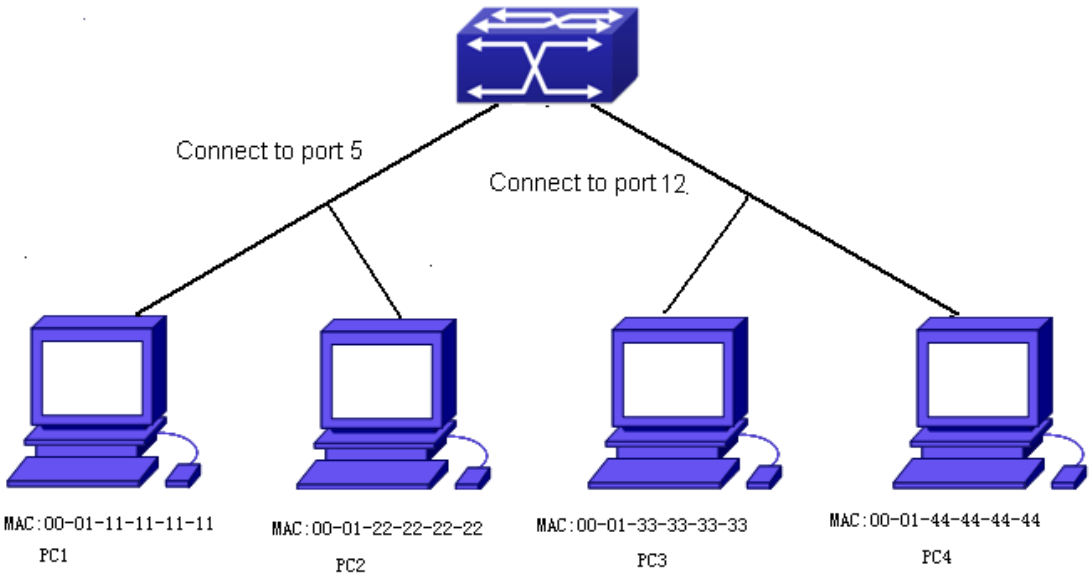


Figure 2-358 MAC Table dynamic learning

The topology of the figure above: 4 PCs connected to switch, where PC1 and PC2 belongs to

a same physical segment (same collision domain), the physical segment connects to port 1/0/5 of switch; PC3 and PC4 belongs to the same physical segment that connects to port 1/0/12 of switch.

The initial MAC table contains no address mapping entries. Take the communication of PC1 and PC3 as an example, the MAC address learning process is as follow:

1. When PC1 sends message to PC3, the switch receives the source MAC address 00-01-11-11-11-11 from this message, the mapping entry of 00-01-11-11-11-11 and port 1/0/5 is added to the switch MAC table.
2. At the same time, the switch learns the message is destined to 00-01-33-33-33-33, as the MAC table contains only a mapping entry of MAC address 00-01-11-11-11-11 and port1/0/5, and no port mapping for 00-01-33-33-33-33 present, the switch broadcast this message to all the ports in the switch (assuming all ports belong to the default VLAN1).
3. PC3 and PC4 on port 1/0/12 receive the message sent by PC1, but PC4 will not reply, as the destination MAC address is 00-01-33-33-33-33, only PC3 will reply to PC1. When port 1/0/12 receives the message sent by PC3, a mapping entry for MAC address 00-01-33-33-33-33 and port 1/0/12 is added to the MAC table.
4. Now the MAC table has two dynamic entries, MAC address 00-01-11-11-11-11 - port 1/0/5 and 00-01-33-33-33-33 -port1/0/12.
5. After the communication between PC1 and PC3, the switch does not receive any message sent from PC1 and PC3. And the MAC address mapping entries in the MAC table are deleted in 300 to 2*300 seconds (ie, in single to double aging time). The 300 seconds here is the default aging time for MAC address entry in switch. Aging time can be modified in switch.

2.22.1.2 Forward or Filter

The switch will forward or filter received data frames according to the MAC table. Take the above figure as an example, assuming switch have learnt the MAC address of PC1 and PC3, and the user manually configured the mapping relationship for PC2 and PC4 to ports. The MAC table of switch will be:

MAC Address	Port number	Entry added by
00-01-11-11-11-11	1/0/5	Dynamic learning
00-01-22-22-22-22	1/0/5	Static configuration
00-01-33-33-33-33	1/0/12	Dynamic learning
00-01-44-44-44-44	1/0/12	Static configuration

1. Forward data according to the MAC table

If PC1 sends a message to PC3, the switch will forward the data received on port 1/0/5 from port1/0/12.

2. Filter data according to the MAC table

If PC1 sends a message to PC2, the switch, on checking the MAC table, will find PC2 and PC1 are in the same physical segment and filter the message (i.e. drop this message).

Three types of frames can be forwarded by the switch:

- ☞ Broadcast frame

- ☞ Multicast frame
- ☞ Unicast frame

The following describes how the switch deals with all the three types of frames:

1. Broadcast frame: The switch can segregate collision domains but not broadcast domains. If no VLAN is set, all devices connected to the switch are in the same broadcast domain. When the switch receives a broadcast frame, it forwards the frame in all ports. When VLANs are configured in the switch, the MAC table will be adapted accordingly to add VLAN information. In this case, the switch will not forward the received broadcast frames in all ports, but forward the frames in all ports in the same VLAN.
2. Multicast frame: For the unknown multicast, the switch will broadcast it in the same vlan, but the switch only forwards the multicast frames to the multicast group's port if IGMP Snooping function or the static multicast group has been configured.
3. Unicast frame: When no VLAN is configured, if the destination MAC addresses are in the switch MAC table, the switch will directly forward the frames to the associated ports; when the destination MAC address in a unicast frame is not found in the MAC table, the switch will broadcast the unicast frame. When VLANs are configured, the switch will forward unicast frame within the same VLAN. If the destination MAC address is found in the MAC table but belonging to different VLANs, the switch can only broadcast the unicast frame in the VLAN it belongs to.

2.22.2 Mac Address Table Configuration Task List

1. Configure the MAC address aging-time
2. Configure static MAC forwarding or filter entry
3. Clear dynamic address table
4. Configure MAC learning through CPU control
5. Configure the hash collision mac table

1. Configure the MAC aging-time

Command	Explanation
Global Mode	
mac-address-table aging-time <*0*|*aging-time*> **no mac-address-table aging-time**	Configure the MAC address aging-time.

Configure static MAC forwarding or filter entry

Command	Explanation
Global Mode	

mac-address-table {static | static-multicast | blackhole} address *<mac-addr>* vlan *<vlan-id >* [interface [ethernet | portchannel] *<interface-name>*] | [source|destination|both] **no mac-address-table {static | static-multicast | blackhole | dynamic} *[address* <mac-addr>] *[vlan* <vlan-id>] *[interface* [ethernet | portchannel] *<interface-name>*]**	Configure static MAC entries, static multicast MAC entries, filter address entires.

2. Clear dynamic address table

Command	Explanation
Admin Mode	
clear mac-address-table dynamic [address *<mac-addr>*] [vlan *<vlan-id>*] [interface [ethernet | portchannel] *<interface-name>*]	Clear the dynamic address table.

4. **Configure MAC learning through CPU control**

Command	Explanation
Global Mode	
mac-address-learning cpu-control **no mac-address-learning cpu-control**	Enable MAC learning through CPU control, the no command restores that the chip automatically learn MAC address.

5. **Configure bucket size for MAC address table**

Command	Explanation
Global Mode	
show collision-mac-address-table	Show the hash collision mac table.
Admin Mode	
clear collision-mac-address-table	Clear the hash collision mac table.

2.22.3 Typical Configuration Examples

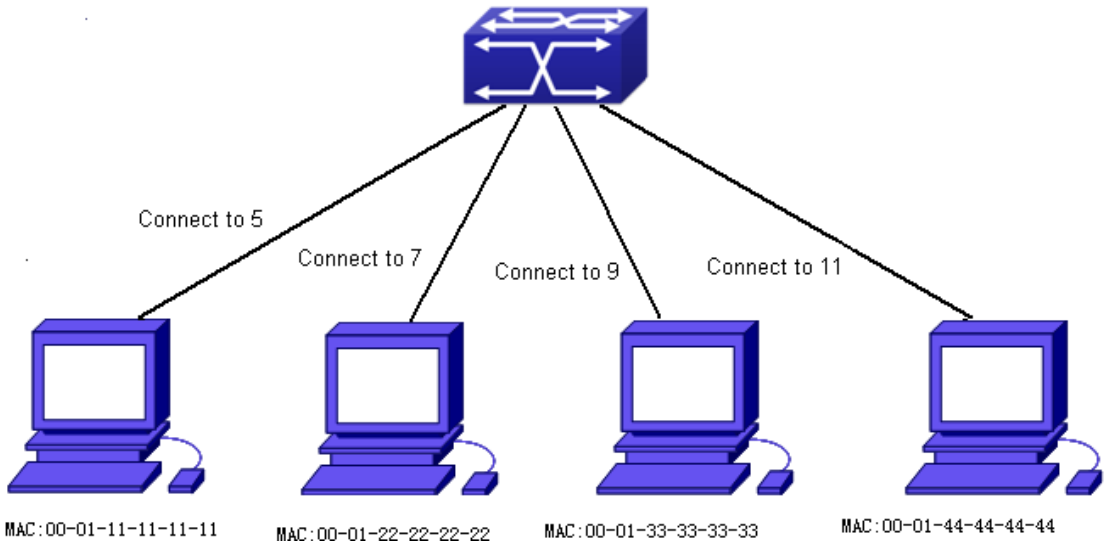


Figure 2-369 MAC Table typical configuration example

Scenario:

Four PCs as shown in the above figure connect to port 1/0/5, 1/0/7, 1/0/9, 1/0/11 of switch, all the four PCs belong to the default VLAN1. As required by the network environment, dynamic learning is enabled. PC1 holds sensitive data and can not be accessed by any other PC that is in another physical segment; PC2 and PC3 have static mapping set to port 1/0/7 and port 1/0/9, respectively.

The configuration steps are listed below:

1. Set the MAC address 00-01-11-11-11-11 of PC1 as a filter address.

Switch(config)#mac-address-table static 00-01-11-11-11-11 discard vlan 1.

2.Set the static mapping relationship for PC2 and PC3 to port 1/0/7 and port 1/0/9, respectively.

Switch(config)#mac-address-table static address 00-01-22-22-22-22 vlan 1 interface ethernet 1/0/7

Switch(config)#mac-address-table static address 00-01-33-33-33-33 vlan 1 interface ethernet 1/0/9

2.22.4 MAC Table Troubleshooting

Using the show mac-address-table command, a port is found to be failed to learn the MAC of a device connected to it. Possible reasons:

- The connected cable is broken.
- Spanning Tree is enabled and the port is in “discarding” status; or the device is just connected to the port and Spanning Tree is still under calculation, wait until the Spanning Tree calculation finishes, and the port will learn the MAC address.
- If not the problems mentioned above, please check for the switch portand contact technical support for solution.

2.23 MAC Notification

2.23.1 Introduction to MAC Notification

MAC Notification function depends on the notification. Add or remove the MAC address, namely, when the device is added or removed, it will notify administrator about the changing by the trap function of snmp.

2.23.2 MAC Notification Configuration

Mac notification configuration task list:
1. Configure the global snmp MAC notification
2. Configure the global MAC notification
3. Configure the interval for sending MAC notification
4. Configure the size of history table
5. Configure the trap type of MAC notification supported by the port
6. Show the configuration and the data of MAC notification
7. Clear the statistics of MAC notification trap

1. Configure the global snmp MAC notification

Command	Explanation
Global mode	
snmp-server enable traps mac-notification **no snmp-server enable traps mac-notification**	Configure or cancel the global snmp MAC notification.

2. Configure the global MAC notification

Command	Explanation
Global mode	
mac-address-table notification **no mac-address-table notification**	Configure or cancel the global MAC notification.

3. Configure the interval for sending MAC notification

Command	Explanation
Global mode	
mac-address-table notification interval <0-86400> **no mac-address-table notification interval**	Configure the interval for sending the MAC address notification, the no command restores the default interval.

4. Configure the size of history table

Command	Explanation

Global mode	
mac-address-table notification history-size <0-500> **no mac-address-table notification history-size**	Configure the history table size, the no command restores the default value.

5. Configure the trap type of MAC notification supported by the port

Command	Explanation
Port mode	
mac-notification {added | both | removed} **no mac-notification**	Configure or cancel the trap type of MAC notification supported by the port.

6. Show the configuration and the data of MAC notification

Command	Explanation
Admin mode	
show mac-notification summary	Show the configuration and the data of MAC notification.

7. Clear the statistics of MAC notification trap

Command	Explanation
Admin mode	
clear mac-notification statistics	Clear the statistics of MAC notification trap.

2.23.3 MAC Notification Example

IP address of network management station (NMS) is 1.1.1.5, IP address of Agent is 1.1.1.9. NMS will receive Trap message from Agent. (Note: NMS may set the authentication to the community character string of trap, suppose the community character string as usertrap)
Configuration procedure in the following:
Switch(config)#snmp-server enable
Switch(config)#snmp-server enable traps mac-notification
Switch(config)# mac-address-table notification
Switch(config)# mac-address-table notification interval 5
Switch(config)# mac-address-table notification history-size 100
Switch(Config-If-Ethernet1/0/4)# mac-notification both

2.23.4 MAC Notification Troubleshooting

Check whether trap message is sent successfully by show command and debug command of snmp.

Chapter 3 IP services Configuration

3.1 Layer 3 Interface

3.1.1 Introduction to Layer 3 Interface

Layer 3 interface can be created on switch. The Layer 3 interface is not a physical interface but a virtual interface. Layer 3 interface is built on VLANs. The Layer 3 interface can contain one or more layer 2 ports which belong to the same VLAN, or contain no layer 2 ports. At least one of the Layer 2 ports contained in Layer 3 interface should be in UP state for Layer 3 interface in UP state, otherwise, Layer 3 interface will be in DOWN state. All layer 3 interfaces in the switch use the same MAC address by default, this address is selected from the reserved MAC address while creating Layer 3 interface. The Layer 3 interface is the base for layer 3 protocols. The switch can use the IP addresses set in the layer 3 interfaces to communicate with the other devices via IP. The switch can forward IP packets between different Layer 3 interfaces. Loopback interface belongs to Layer 3 interface.

3.1.2 Layer 3 Interface Configuration Task List

Layer 3 Interface Configuration Task List:

1. Create Layer 3 interface
2. Bandwidth for Layer 3 Interface configuration
3. Configure VLAN interface description
4. Open or close the VLAN interface
5. VRF configuration

(1) Create VRF instance and enter VPN view

(2) Configure RD of VRF instance (optional)

(3) Configure RT of VRF instance (optional)

(4) Configure the relation between VRF instance and the interface

1. Create Layer 3 Interface

Command	Explanation
Global Mode	
interface vlan *<vlan-id>* **no interface vlan *<vlan-id>***	Creates a VLAN interface (VLAN interface is a Layer 3 interface); the no command deletes the VLAN interface (Layer 3 interface) created in the switch.
interface loopback *<loopback-id>* ***no*** **interface loopback *<loopback-id>***	Creates a Loopback interface then enter the loopback Port Mode; the no command deletes the Loopback interface created in the switch.

2. Bandwidth for Layer 3 Interface configuration

Command	Explanation
VLAN Interface Mode	
bandwidth <bandwidth> ***no* bandwidth**	Configure the bandwidth for Layer 3 Interface. The no command recovery the default value.

3. Configure VLAN interface description

Command	Explanation
VLAN Interface Mode	
description *<text>* **no description**	Configure the description information of VLAN interface. The no command will cancel the description information of VLAN interface.

4. Open or close the vlan interface

Command	Explanation
VLAN Interface Mode	
shutdown **no shutdown**	Open or close the vlan interface.

5. VRF configuration

(1) Create VRF instance and enter VPN view

(2) Configure RD of VRF instance (optional)

(3) Configure RT of VRF instance (optional)

(4) Configure the relation between VRF instance and the interface

Command	Explanation
Global Mode	
ip vrf <vrf-name> **no ip vrf <vrf-name>**	Create VRF instance; VRF instance is not created by default.
VRF Mode	
rd <ASN:nn_or_IP-address:nn>	Configure RD of VRF instance. RD is not created by default.
route-target {import | export | both} <rt-value> **no route-target {import | export | both} <rt-value>**	Configure RT of VRF instance
Interface Mode	
ip vrf forwarding <vrf-name> **no ip vrf forwarding <vrf-name>**	Configure the relation between VRF instance and the interface.
ip address <ip-address> <mask> **no ip address <ip-address> <mask>**	Configure the private IP address of direct link interface.

3.2 Network Management Port

3.2.1 Introduce to IP Network Management Port

The Network Management Port is located above the Console interface on the master management cards, marked as “Ethernet”, the software configuration name for this port is “Ethernet0”. The user can use the command “interface Ethernet 0” to access the network port configuration mode. The user can use programs such as Telnet, Web management and FTP to manage the switch through an Ethernet cable connected to the network management port.

3.2.2 Network Management Port Configuration

Network Management Port Configuration Task List:

1. Enter the network management port configuration mode
2. Configure the properties for the network management ports

（1） Enable/Disable ports

（2） Configuring port IP Address

1. Enter the network management port configuration mode

Command	Explanation
Global Mode	
interface ethernet *<num>*	Enters the network management port configuration mode.

2. Configure the properties for the network management port

Command	Explanation
Network Management Port Configuration	
shutdown **no shutdown**	Enables/Disables network management port.
ip address *<ip-address>* *<mask>* **no ip address [*<ip-address>* *<mask>*]**	Configures or cancels the IP address for network management port.

3.3 IP Configuration

3.3.1 Introduction to IPv4, IPv6

IPv4 is the current version of global universal Internet protocol. The practice has proved that IPv4 is simple, flexible, open, stable, strong and easy to implement while collaborating well with various protocols of upper and lower layers. Although IPv4 almost has not been changed since it

was established in 1980's, it has kept growing to the current global scale with the promotion of Internet. However, as Internet infrastructure and Internet application services continue boosting, IPv4 has shown its deficiency when facing the present scale and complexity of Internet.

IPv6 refers to the sixth version of Internet protocol which is the next generation Internet protocol designed by IETF to replace the current Internet protocol version 4 (IPv4). IPv6 was specially developed to make up the shortages of IPv4 addresses so that Internet can develop further.

The most important problem IPv6 has solved is to add the amount of IP addresses. IPv4 addresses have nearly run out, whereas the amount of Internet users has been increasing in geometric series. With the greatly and continuously boosting of Internet services and application devices (Home and Small Office Network, IP phone and Wireless Service Information Terminal which make use of Internet,) which require IP addresses, the supply of IP addresses turns out to be more and more tense. People have been working on the problem of shortage of IPv4 addresses for a long time by introducing various technologies to prolong the lifespan of existing IPv4 infrastructure, including Network Address Translation(NAT for short), and Classless Inter-Domain Routing(CIDR for short), etc.

Although the combination of CIDR, NAT and private addressing has temporarily mitigated the problem of IPv4 address space shortage, NAT technology has disrupted the end-to-end model which is the original intention of IP design by making it necessary for router devices that serve as network intermediate nodes to maintain every connection status which increases network delay greatly and decreases network performance. Moreover, the translation of network data packet addresses baffles the end-to-end network security check, IPSec authentication header is such an example.

Therefore, in order to solve all kinds of problems existing in IPv4 comprehensively, the next generation Internet Protocol IPv6 designed by IETF has become the only feasible solution at present.

First of all, the 128 bits addressing scheme of IPv6 Protocol can guarantee to provide enough globally unique IP addresses for global IP network nodes in the range of time and space. Moreover, besides increasing address space, IPv6 also enhanced many other essential designs of IPv4.

Hierarchical addressing scheme facilitates Route Aggregation, effectively reduces route table entries and enhances the efficiency and expansibility of routing and data packet processing.

The header design of IPv6 is more efficient compared with IPv4. It has less data fields and takes out header checksum, thus expedites the processing speed of basic IPv6 header. In IPv6 header, fragment field can be shown as an optional extended field, so that data packets fragmentation process won't be done in router forwarding process, and Path MTU Discovery Mechanism collaborates with data packet source which enhances the processing efficiency of router.

Address automatic configuration and plug-and-play is supported. Large amounts of hosts can find network routers easily by address automatic configuration function of IPv6 while obtaining a globally unique IPv6 address automatically as well which makes the devices using IPv6 Internet plug-and-play. Automatic address configuration function also makes the readdressing of existing network easier and more convenient, and it is more convenient for network operators to manage the transformation from one provider to another.

Support IPSec. IPSec is optional in IPv4, but required in IPv6 Protocol. IPv6 provides security extended header, which provides end-to-end security services such as access control, confidentiality and data integrity, consequently making the implement of encryption, validation and Virtual Private Network easier.

Enhance the support for Mobile IP and mobile calculating devices. The Mobile IP Protocol defined in IETF standard makes mobile devices movable without cutting the existing connection, which is a network function getting more and more important. Unlike IPv4, the mobility of IPv6 is from embedded automatic configuration to get transmission address (Care-Of-Address); therefore it doesn't need Foreign Agent. Furthermore, this kind of binding process enables Correspondent Node communicate with Mobile Node directly, thereby avoids the extra system cost caused by triangle routing choice required in IPv4.

Avoid the use of Network Address Translation. The purpose of the introduction of NAT mechanism is to share and reuse same address space among different network segments. This mechanism mitigates the problem of the shortage of IPv4 address temporally; meanwhile it adds the burden of address translation process for network device and application. Since the address space of IPv6 has increased greatly, address translation becomes unnecessary, thus the problems and system cost caused by NAT deployment are solved naturally.

Support extensively deployed Routing Protocol. IPv6 has kept and extended the supports for existing Internal Gateway Protocols (IGP for short), and Exterior Gateway Protocols (EGP for short). For example, IPv6 Routing Protocol such as RIPng, OSPFv3, IS-ISv6 and MBGP4+, etc.

Multicast addresses increased and the support for multicast has enhanced. By dealing with IPv4 broadcast functions such as Router Discovery and Router Query, IPv6 multicast has completely replaced IPv4 broadcast in the sense of function. Multicast not only saves network bandwidth, but enhances network efficiency as well.

3.3.2 IP Configuration

Layer 3 interface can be configured as IPv4 interface, IPv6 interface.

3.3.2.1 IPv4 Address Configuration

IPv4 address configuration task list:

1. Configure the IPv4 address of three-layer interface

1. Configure the IPv4 address of three-layer interface

Command	Explanation
VLAN Interface Configuration Mode	
ip address *<ip-address> <mask>* [secondary] **no ip address [*<ip-address> <mask>*]**	Configure IP address of VLAN interface; the **no ip address [*<ip-address> <mask>*]** command cancels IP address of VLAN interface.

3.3.2.2 IPv6 Address Configuration

The configuration Task List of IPv6 is as follows:

1. IPv6 basic configuration
   (1) Configure interface IPv6 address
   (2) Configure IPv6 static routing
   2. IPv6 Neighbor Discovery Configuration
   (1) Configure DAD neighbor solicitation message number
(2) Configure send neighbor solicitation message interval
   (3) Enable and disable router advertisement
   (4) Configure router lifespan
   (5) Configure router advertisement minimum interval
   (6) Configure router advertisement maximum interval
   (7) Configure prefix advertisement parameters
   (8) Configure static IPv6 neighbor entries
   (9) Delete all entries in IPv6 neighbor table
   (10) Set the hoplimit of sending router advertisement
   (11) Set the mtu of sending router advertisement
   (12) Set the reachable-time of sending router advertisement
   (13) Set the retrans-timer of sending router advertisement
   (14) Set the flag representing whether information other than the address information will be obtained via DHCPv6
   (15) Set the flag representing whether the address information will be obtained via DHCPv6
3. IPv6 Tunnel configuration
   (1) Create/Delete Tunnel
   (2) Configure tunnel description
   (3) Configure Tunnel Source
   (4) Configure Tunnel Destination
   (5) Configure Tunnel Next-Hop
   (6) Configure Tunnel Mode
   (7) Configure Tunnel Routing

1. **IPv6 Basic Configuration**

(1) Configure interface IPv6 address

Command	Explanation
Interface Configuration Mode	
ipv6 address ***<ipv6-address/prefix-length>*** **[eui-64]** **no ipv6 address** ***<ipv6-address/prefix-length>***	Configure IPv6 address, including aggregatable global unicast addresses, site-local addresses and link-local addresses. The **no ipv6 address** ***<ipv6-address/prefix-length>*** command cancels IPv6 address.

(2) Set IPv6 Static Routing

Command	Explanation

Global mode	
ipv6 route <*ipv6-prefix/prefix-length*> {<nexthop-ipv6-address>|<interface-type interface-number> | {<nexthop-ipv6-address> <interface-type interface-number>}} [distance] **no ipv6 route <*ipv6-prefix/prefix-length*> {<nexthop-ipv6-address>|<interface-type interface-number> |{<nexthop-ipv6-address> <interface-type interface-number>}} [distance]**	Configure IPv6 static routing. The no command cancels IPv6 static routing.

2. IPv6 Neighbor Discovery Configuration

(1) Configure DAD Neighbor solicitation Message number

Command	Explanation
Interface Configuration Mode	
ipv6 nd dad attempts <value> **no ipv6 nd dad attempts**	Set the neighbor query message number sent in sequence when the interface makes duplicate address detection. The no command resumes default value (1).

(2) Configure Send Neighbor solicitation Message Interval

Command	Explanation
Interface Configuration Mode	

ipv6 nd ns-interval <seconds> **no ipv6 nd ns-interval**	Set the interval of the interface to send neighbor query message. The NO command resumes default value (1 second).

(3) Enable and disable router advertisement

Command	Explanation
Interface Configuration Mode	
ipv6 nd suppress-ra **no ipv6 nd suppress-ra**	Forbid IPv6 Router Advertisement. The NO command enables IPv6 router advertisement.

(4) Configure Router Lifespan

Command	Explanation
Interface Configuration Mode	

ipv6 nd ra-lifetime <seconds> **no ipv6 nd ra-lifetime**	Configure Router advertisement Lifespan. The NO command resumes default value (1800 seconds).

(5) Configure router advertisement Minimum Interval

Command	Description
Interface Configuration Mode	
ipv6 nd min-ra-interval <seconds> **no ipv6 nd min-ra-interval**	Configure the minimum interval for router advertisement. The NO command resumes default value (200 seconds).

(6) Configure router advertisement Maximum Interval

Command	Explanation
Interface Configuration Mode	

ipv6 nd max-ra-interval <seconds> **no ipv6 nd max-ra-interval**	Configure the maximum interval for router advertisement. The NO command resumes default value (600 seconds).

(7) Configure prefix advertisement parameters

Command	Explanation
Interface Configuration Mode	
ipv6 nd prefix **<ipv6-address/prefix-length>** ***<valid-lifetime> <preferred-lifetime>*** **[off-link] [no-autoconfig]** **no ipv6 nd prefix** ***<ipv6-address/prefix-length>*** ***<valid-lifetime> <preferred-lifetime>*** **[off-link] [no-autoconfig]**	Configure the address prefix and advertisement parameters of router. The NO command cancels the address prefix of routing advertisement.

(8) Configure static IPv6 neighbor Entries

Command	Explanation
Interface Configuration Mode	

ipv6 neighbor <ipv6-address> <hardware-address> interface *<interface-type interface-name>*	Set static neighbor table entries, including neighbor IPv6 address, MAC address and two-layer port.
no ipv6 neighbor <ipv6-address>	Delete neighbor table entries.

(9) Delete all entries in IPv6 neighbor table

Command	Explanation
Admin Mode	
clear ipv6 neighbors	Clear all static neighbor table entries.

(10) Set the hoplimit of sending router advertisement

Command	Explanation
Interface Configuration Mode	
ipv6 nd ra-hoplimit <value>	Set the hoplimit of sending router advertisement.

(11) Set the mtu of sending router advertisement

Command	Explanation
Interface Configuration Mode	
ipv6 nd ra-mtu <value>	Set the mtu of sending router advertisement.

(12) Set the reachable-time of sending router advertisement

Command	Explanation
Interface Configuration Mode	
ipv6 nd reachable-time <seconds>	Set the reachable-time of sending router advertisement.

(13) Set the retrans-timer of sending router advertisement

Command	Explanation
Interface Configuration Mode	
ipv6 nd retrans-timer <seconds>	Set the retrans-timer of sending router advertisement.

(14) Set the flag representing whether information other than the address information will be obtained via DHCPv6.

Command	Explanation
Interface Configuration Mode	
ipv6 nd other-config-flag	Set the flag representing whether information other than the address information will be obtained via DHCPv6.

(15) Set the flag representing whether the address information will be obtained via DHCPv6

Command	Explanation
Interface Configuration Mode	
ipv6 nd managed-config-flag	Set the flag representing whether the address information will be obtained via DHCPv6.

3. IPv6 Tunnel Configuration

(1) Add/Delete tunnel

Command	Explanation
Global mode	
interface tunnel *<tnl-id>* **no interface tunnel *<tnl-id>***	Create a tunnel. The NO command deletes a tunnel.

(2) Configure tunnel description

Command	Explanation
Tunnel Configuration Mode	
description *<desc>* **no description**	Configure tunnel description. The NO command deletes the tunnel description.

(3) Configure tunnel source

Command	Explanation
Tunnel Configuration Mode	

tunnel source { *<ipv4-address>* / *<ipv6-address>*/ *<interface-name>* } **no tunnel source**	Configure tunnel source end IPv4/IPv6 address. The NO command deletes the IPv4/IPv6 address of tunnel source end.

(4) Configure Tunnel Destination

Command	Explanation
Tunnel Configuration Mode	
tunnel destination {<ipv4-address> | <ipv6-address>} **no tunnel destination**	Configure tunnel destination end IPv4/IPv6 address. The NO command deletes the IPv4/IPv6 address of tunnel destination end.

(5) Configure Tunnel Next-Hop

Command	Explanation
Tunnel Configuration Mode	
tunnel nexthop <ipv4-address> **no tunnel nexthop**	Configure tunnel next-hop IPv4 address. The NO command deletes the IPv4 address of tunnel next-hop end.

(6) Configure Tunnel Mode

Command	Explanation
Tunnel Configuration Mode	
tunnel mode [[gre] | ipv6ip [6to4 | isatap]] **no tunnel mode**	Configure tunnel mode. The NO command clears tunnel mode.

(7) Configure Tunnel Routing

Command	Explanation
Global mode	
ipv6 route <ipv6-address/prefix-length> {<interface-type interface-number> | tunnel <tnl-id>} **no ipv6 route <ipv6-address/prefix-length> {<interface-type interface-number> | tunnel <tnl-id>}**	Configure tunnel routing. The NO command clears tunnel routing.

3.3.3 IP Configuration Examples

3.3.3.1 Configuration Examples of IPv4

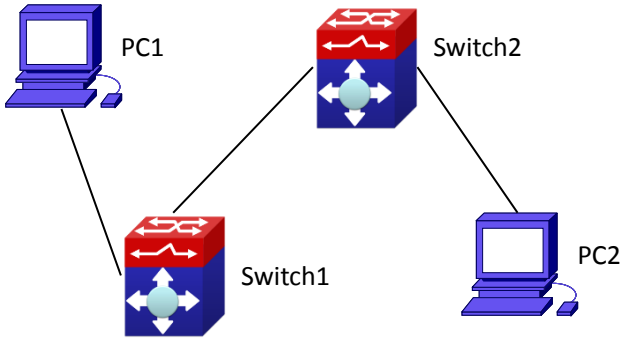


Figure 3-1 IPv4 configuration example

The user's configuration requirements are: Configure IP address of different network segments on Switch1 and Switch2, configure static routing and validate accessibility using ping function.

Configuration Description:

Configure two VLANs on Switch1, namely, VLAN1 and VLAN2.

Configure IPv4 address 192.168.1.1 255.255.255.0 in VLAN1 of Switch1, and configure IPv4 address 192.168.2.1 255.255.255.0 in VLAN2.

Configure two VLANs on Switch2, respectively VLAN2 and VLAN3.

Configure IPv4 address 192.168.2.2 255.255.255.0 in VLAN2 of Switch2, and configure IPv4 address 192.168.3.1 255.255.255.0 in VLAN3.

The IPv4 address of PC1 is 192.168.1.100 255.255.255.0, and the IPv4 address of PC2 is 192.168.3.100 255.255.255.0.

1. Configure static routing 192.168.3.0/24 on Switch1, and configure static routing 192.168.1.0/24 on Switch2.
2. Ping each other among PCs.
3. Note: First make sure PC1 and Switch1 can access each other by ping, and PC2 and Switch2 can access each other by ping.
4. The configuration procedure is as follows:
5. Switch1(config)#interface vlan 1
6. Switch1(Config-if-Vlan1)#ip address 192.168.1.1 255.255.255.0
7. Switch1(config)#interface vlan 2
8. Switch1(Config-if-Vlan2)#ip address 192.168.2.1 255.255.255.0
9. Switch1(Config-if-Vlan2)#exit
10. Switch1(config)#ip route 192.168.3.0 255.255.255.0 192.168.2.2
11.
12. Switch2(config)#interface vlan 2
13. Switch2(Config-if-Vlan2)#ip address 192.168.2.2 255.255.255.0
14. Switch2(config)#interface vlan 3
15. Switch2(Config-if-Vlan3)#ip address 192.168.3.1 255.255.255.0
16. Switch2(Config-if-Vlan3)#exit
17. Switch2(config)#ip route 192.168.1.0 255.255.255.0 192.168.2.1

3.3.3.2 Configuration Examples of IPv6

Example 1:

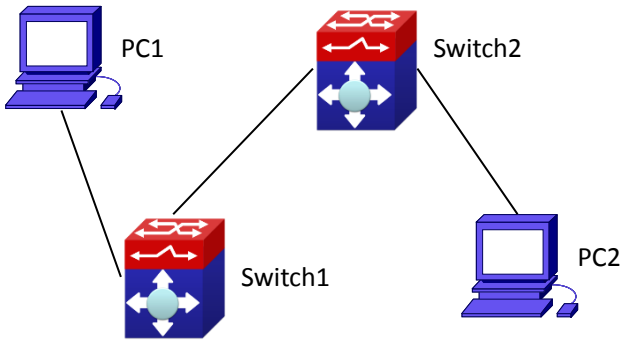


Figure 3-2 IPv6 configuration example

The user's configuration requirements are: Configure IPv6 address of different network segments on Switch1 and Switch2, configure static routing and validate reachability using ping6 function.

Configuration Description:

Configure two VLANs on Switch1, namely, VLAN1 and VLAN2.

Configure IPv6 address 2001::1/64 in VLAN1 of Switch1, and configure IPv6 address 2002::1/64 in VLAN2.

Configure 2 VLANs on Switch2, namely, VLAN2 and VLAN3.

Configure IPv6 address 2002::2/64 in VLAN2 of Switch2, and configure IPv6 address 2003::1/64 in VLAN3.

1. The IPv6 address of PC1 is 2001::11/64, and the IPv6 address of PC2 is 2003::33/64.
2. Configure static routing 2003:33/64 on Switch1, and configure static routing 2001::11/64 on Switch2.
3. ping6 each other among PCs.
4. Note: First make sure PC1 and Switch1 can access each other by ping, and PC2 and Switch2 can access each other by ping.
5. The configuration procedure is as follows:
6. Switch1(Config)#interface vlan 1
7. Switch1(Config-if-Vlan1)#ipv6 address 2001::1/64
8. Switch1(Config)#interface vlan 2
9. Switch1(Config-if-Vlan2)#ipv6 address 2002::1/64
10. Switch1(Config-if-Vlan2)#exit
11. Switch1(Config)#ipv6 route 2003::33/64 2002::2
12.
13. Switch2(Config)#interface vlan 2
14. Switch2(Config-if-Vlan2)#ipv6 address 2002::2/64

15. Switch2(Config)#interface vlan 3

16. Switch2(Config-if-Vlan3)#ipv6 address 2003::1/64

Switch2(Config-if-Vlan3)#exit

Switch2(Config)#ipv6 route 2001::33/64 2002::1

Switch1#ping6 2003::33

Configuration result:

Switch1#show run

interface Vlan1

ipv6 address 2001::1/64

!

interface Vlan2

ipv6 address 2002::2/64

!

interface Loopback

mtu 3924

!

ipv6 route 2003::/64 2002::2

!

no login

!

end

Switch2#show run

interface Vlan2

ipv6 address 2002::2/64

!

interface Vlan3

ipv6 address 2003::1/64

!

interface Loopback

mtu 3924

!

ipv6 route 2001::/64 2002::1

!

no login

!

End

Example **2:**

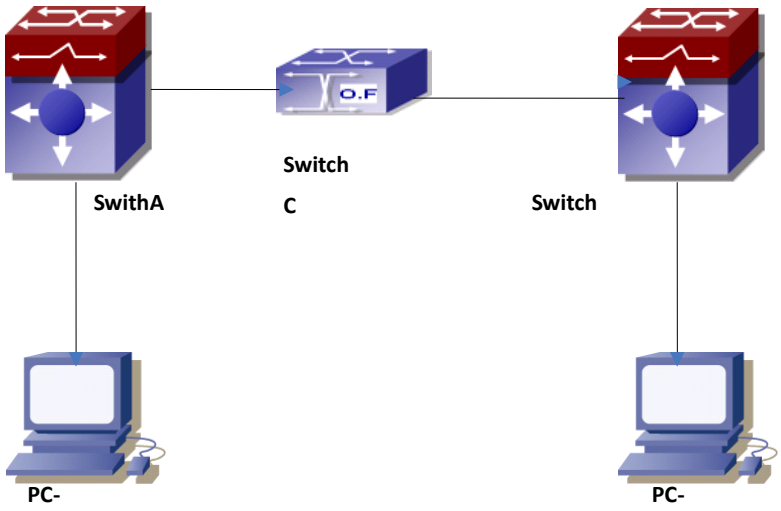


Figure 3-3 IPv6 tunnel

This case is IPv6 tunnel with the following user configuration requirements: SwitchA and SwitchB are tunnel nodes, dual-stack is supported. SwitchC only runs IPv4, PC-A and PC-B communicate.

Configuration Description:

Configure two vlans on SwitchA, namely, VLAN1 and VLAN2. VLAN1 is IPv6 domain, VLAN2 connects to IPv4 domain.

Configure IPv6 address 2002:caca:ca01:2::1/64 in VLAN1 of SwitchA and turn on RA function, configure IPv4 address 202.202.202.1 in VLAN2.

Configure two VLANs on SwitchB, namely, VLAN3 and VLAN4, VLAN4 is IPv6 domain, and VLAN3 connects to IPv4 domain.

Configure IPv6 address 2002:cbcb:cb01:2::1/64 in VLAN4 of SwitchB and turn on RA function, configure IPv4 address 203.203.203.1 on VLAN3.

Configure tunnel on SwitchA, the source IPv4 address of the tunnel is 202.202.202.1, the tunnel routing is ::/0

Configure tunnel on SwitchB, the source IPv4 address of the tunnel is 203.203.203.1, and the tunnel routing is ::/0

Configure two VLANs on SwitchC, namely, VLAN2 and VLAN3. Configure IPv4 address 202.202.202.202 on VLAN2 and configure IPv4 address 203.203.203.203 on VLAN3.

PC-A and PC-B get the prefix of 2002 via SwitchA and SwitchB to configure IPv6 address automatically.

On PC-A, ping IPv6 address of PC-B

The configuration procedure is as follows:

SwitchA(Config-if-Vlan1)#ipv6 address 2002:caca:ca01:2::1/64

SwitchA(Config-if-Vlan1)#no ipv6 nd suppress-ra

SwitchA(Config-if-Vlan1)#interface vlan 2

SwitchA(Config-if-Vlan2)#ipv4 address 202.202.202.1 255.255.255.0

SwitchA(Config-if-Vlan1)#exit

SwitchA(config)# interface tunnel 1

SwitchA(Config-if-Tunnel1)#tunnel source 202.202.202.1
SwitchA(Config-if-Tunnel1)#tunnel destination 203.203.203.1
SwitchA(Config-if-Tunnel1)#tunnel mode ipv6ip
SwitchA(config)#ipv6 route ::/0 tunnel1

SwitchB(Config-if-Vlan4)#ipv6 address 2002:cbcb:cb01::2/64
SwitchB(Config-if-Vlan4)#no ipv6 nd suppress-ra
SwitchB (Config-if-Vlan3)#interface vlan 3
SwitchB (Config-if-Vlan2)#ipv4 address 203.203.203.1 255.255.255.0
SwitchB (Config-if-Vlan1)#exit
SwitchB(config)#interface tunnel 1
SwitchB(Config-if-Tunnel1)#tunnel source 203.203.203.1
SwitchB(Config-if-Tunnel1)#tunnel destination 202.202.202.1
SwitchB(Config-if-Tunnel1)#tunnel mode ipv6ip
SwitchB(config)#ipv6 route ::/0 tunnel1

3.3.4 IPv6 Troubleshooting

The router lifespan configured should not be smaller than the Send Router advertisement Interval. If the connected PC has not obtained IPv6 address, you should check RA announcement switch (the default is turned off).

3.4 IP Forwarding

3.4.1 Introduction to IP Forwarding

Gateway devices can forward IP packets from one subnet to another; such forwarding uses routes to find a path. IP forwarding of switch is done with the participation of hardware, and can achieve wire speed forwarding. In addition, flexible management is provided to adjust and monitor forwarding. Switch supports aggregation algorithm enabling/disabling optimization to adjust generation of network route entry in the switch chip and view statistics for IP forwarding and hardware forwarding chip status.

3.4.2 IP Route Aggregation Configuration Task

IP route aggregation configuration task:
1. Set whether IP route aggregation algorithm with/without optimization should be used

1. Set whether IP route aggregation algorithm with/without optimization should be used

Command	Explanation
Global Mode	
ip fib optimize **no ip fib optimize**	Enables the switch to use optimized IP route aggregation algorithm; the "**no ip fib optimize**" disables the optimized IP route aggregation algorithm.

3.5 URPF

3.5.1 Introduction to URPF

URPF (Unicast Reverse Path Forwarding) introduces the RPF technology applied in multicast to unicast, so to protect the network from the attacks which is based on source address cheat.

When switch receives the packet, it will search the route in the route table using the source address as the destination address which is acquired from the packet. If the found router exit interface does not match the entrance interface acquired from this packet, the switch will consider this packet a fake packet and discard it.

In Source Address Spoofing attacks, attackers will construct a series of messages with fake source addresses. For applications based on IP address verification, such attacks may allow unauthorized users to access the system as some authorized ones, or even the administrator. Even if the response messages can't reach the attackers, they will also damage the targets.

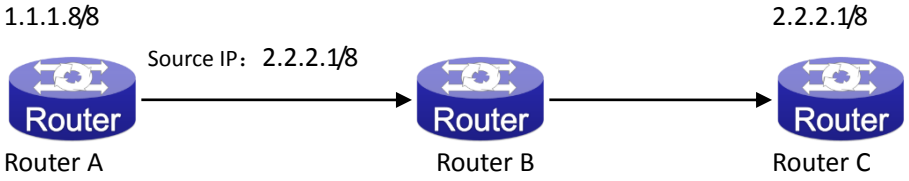


Figure 3-4 URPF application situation

In the above figure, Router A sends requests to the server Router B by faking messages whose source address are 2.2.2.1/8 .In response, Router B will send the messages to the real "2.2.2.1/8". Such illegal messages attack both Router B and Router C. The application of URPF technology in the situation described above can avoid the attacks based on the Source Address Spoofing.

3.5.1.1 IPv6 URPF Operating Mechanism

At present the UPRF relies on the ACL function provided by the switch chips.

Firstly, globally enable the URPF function to monitor the changes in the router table: create

a corresponding URPF permit ACL rule for each router in the router table FIB. In URPF strict mode, the format of ACL rules is: the source address segments of inbound packets + the ingress interface VID of inbound packets. The source address segments of inbound packets are in correspondence with the destination address segments in the FIB router table entries, while the ingress interface VID of inbound packets with the egress interface VID in the FIB router table entries. In URPF loose mode, the format of ACL rules is the source address segments of inbound packets, which are in correspondence with destination address segments in the FIB router table entries.

After enabling URPF on the port: bind the port to RUPF rules, and create the default hardware for DENY ALL rule distribution.

The above operations will guarantee that, when data reach the port, only those match the rules can pass through it with all others dumped.

The present corresponding ACL rule privilege is low, not blocking all kinds of protocol packets; hence, enabling this function will not affect the normal operation of routing protocols of the switch.

3.5.2 URPF Configuration Task Sequence

Enable URPF

2. Enable URPF on port
3. Display and debug URPF relevant information
4.
5. 1. Globally enable URPF

Command	Explanation
Global mode	
urpf enable **no urpf enable**	Globally enable and disable URPF.

2. Enable URPF on port

Command	Explanation
Port mode	
ip urpf enable {loose | strict} {allow-default-route } **no ip urpf enable**	Enable and disable URPF on port.

3. Display and debug URPF relevant information

Command	Explanation
Admin mode	
debug urpf {notice | warn | error|} **no debug urpf {notice | warn | error|}**	Enable the URPF debug function to display error information if failures occur during the installation of URPF rules.
Admin and Config Mode	
show urpf	Display which interfaces have been enabled with URPF function.

show urpf rule ipv4 num interface ethernet IFNAME	Display the number of IPv4 rules bonded to the port.
show urpf rule ipv6 num interface ethernet IFNAME	Display the number of IPv6 rules bonded to the port.
show urpf rule ipv4 interface ethernet IFNAME	Display the details of IPv4 rules bonded to the port.
show urpf rule ipv6 interface ethernet IFNAME	Display the details of IPv6 rules bonded to the port.

3.5.3 URPF Typical Example

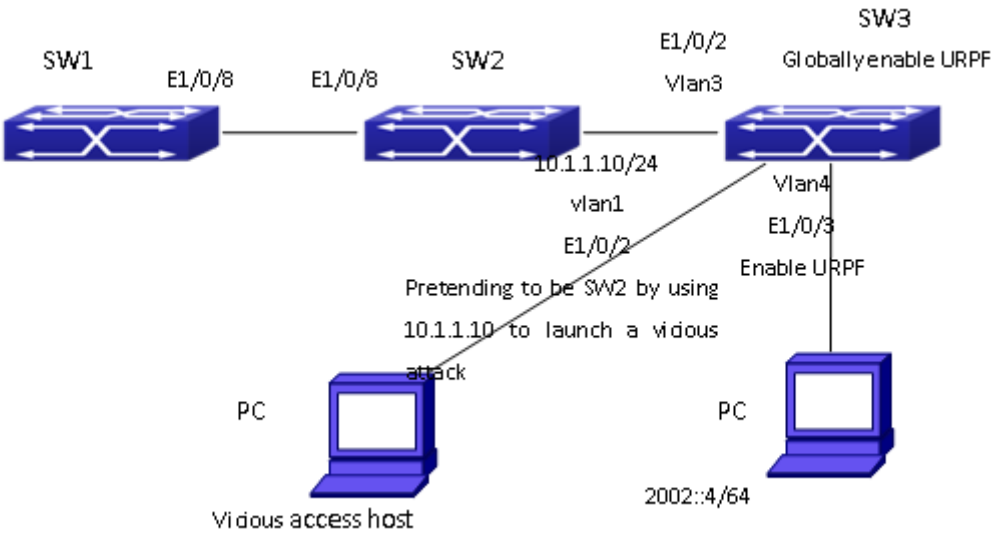


Figure 3-5 URPF Typical Example

In the network, topology shown in the graph above, IP URPF function is enabled on SW3. When there is someone in the network pretending to be someone else by using his IP address to launch a vicious attack, the switch will drop all the attacking messages directly through the hardware FFP function.

Enable the URPF function in SW3 Ethernet1/0/3.

SW3 configuration task sequence:

Switch3#config

Switch3(config)#urpf enable

Switch3(config)#interface ethernet 1/0/3

Switch3(Config-If-Ethernet1/0/3)#ip urpf enable strict

3.5.4 URPF Troubleshooting

Proper operation of the URPF protocol depends greatly on whether the corresponding URPF rules

can be applied correctly. If after the URPF configuration is done and the function does not meet the expectation:

- ☞ Check if the switch has been configured with the rules conflicting with URPF (URPF priority is lower than ACL), the ACL rules will validate if confliction exits.
- ☞ Check whether there is a relative route in the FIB table. Only when one is found, can the ACL rules be distributed to the port.
- ☞ Check if the hardware ACL performance is full which lead to the newly generated route can not be applied with ACL rules.

☞ If all configurations are normal but URPF still can't operate as expected, please enable the URPF debug function and use the

3.6 ARP

3.6.1 Introduction to ARP

ARP (Address Resolution Protocol) is mainly used to resolve IP address to Ethernet MAC address. Switch supports both dynamic ARP and static ARP configuration.Furthermore, switch supports the configuration of proxy ARP for some applications. For instance, when an ARP request is received on the port, requesting an IP address in the same IP segment of the port but not the same physical network, if the port has enabled proxy ARP, the port would reply to the ARP with its own MAC address and forward the actual packets received. Enabling proxy ARP allows machines physically separated but of the same IP segment ignores the physical separation and communicate via proxy ARP interface as if in the same physical network.

3.6.2 ARP Configuration Task List

ARP Configuration Task List:

1. Configure static ARP
2. Configure proxy ARP
3. Clear dynamic ARP
4. Select hash arithmetic
5. Clear the statistic information of ARP messages

1. Configure static ARP

Command	Explanation
VLAN Interface Mode	
arp <*ip_address*> <*mac_address*> {interface [ethernet] <portName>} **no arp <ip_address>**	Configures a static ARP entry; the no command deletes a ARP entry of the specified IP address.

2. Configure proxy ARP

Command	Explanation
VLAN Interface Mode	
ip proxy-arp **no ip proxy-arp**	Enables the proxy ARP function for Ethernet ports: the no command disables the proxy ARP.

3. Clear dynamic ARP

Command	Explanation
Admin mode	
clear arp-cache	Clear the dynamic ARP learnt by the switch.

4. Select hash arithmetic

Command	Explanation
Global mode	
l3 hashselect **[<crc16l|crc16u|crc32l|crc32u|lsb>]**	Set the hash arithmetic of the layer 3 table. This command refers to ARP table list storage in the hardware, the implement need to guide by the technique specialist. The detail information please refer to the interrelated Command Guide.

5. Clear the statistic information of ARP message

Command	Explanation
Admin mode	
clear arp traffic	Clear the statistic information of ARP messages of the switch.

3.6.3 ARP Troubleshooting

If ping from the switch to directly connected network devices fails, the following can be used to check the possible cause and create a solution.

- ☞ Check whether the corresponding ARP has been learned by the switch.
- ☞ If ARP has not been learned, then enabled ARP debugging information and view the sending/receiving condition of ARP packets.
- ☞ Defective cable is a common cause of ARP problems and may disable ARP learning.

3.7 station movement

3.7.1 Introduction to l3 station movement

When arp/nd switch over the port in normal condition, learn the port information of arp/nd entry again according to arp/nd packets. If PC or other network nodes switch over the port, non-security switchover (ARP packets are not sent or received) does not process to learn again. New l3 station movement is used to satisfy arp/nd switchover in specific condition. When MAC switch over the port, it is considered to be security switchover, any network packets (src mac is the network node that process switchover) received from new port spring arp/nd switchover, learn arp/nd to new port.

3.7.2 l3 station movement Configuration Task List

1. Enable/disable l3-station-move

Command	Explanation
Global Mode	
l3-station-move **no l3-station-move**	Enable or disable l3 station move.

3.8 ARP Scanning Prevention

3.8.1 Introduction to ARP Scanning Prevention Function

ARP scanning is a common method of network attack. In order to detect all the active hosts in a network segment, the attack source will broadcast lots of ARP messages in the segment, which will take up a large part of the bandwidth of the network. It might even do large-traffic-attack in the network via fake ARP messages to collapse of the network by exhausting the bandwidth. Usually ARP scanning is just a preface of other more dangerous attack methods, such as automatic virus infection or the ensuing port scanning, vulnerability scanning aiming at stealing information, distorted message attack, and DOS attack, etc.

Since ARP scanning threatens the security and stability of the network with great danger, so it is very significant to prevent it. Switch provides a complete resolution to prevent ARP scanning: if there is any host or port with ARP scanning features is found in the segment, the switch will cut off the attack source to ensure the security of the network.

There are two methods to prevent ARP scanning: port-based and IP-based. The port-based ARP scanning will count the number to ARP messages received from a port in a certain time range, if the number is larger than a preset threshold, this port will be “down”. The IP-based ARP scanning rate-limiting and isolate two levels threshold, when it above level-1 threshold (the limited threshold), the hardware transmits the ARP packet (including ARP request and reply) of this host normally, and only limit the CPU rate. And produce trap warning to notify administrator that there may be attacked; when packets rate is level-2 threshold (isolation threshold), it will take action, record log and produce trap warning. The level-1 limited threshold and level-2 isolate threshold will be open when enable IP-based ARP scanning in global mode, level-1 threshold will take effect until it lower than level-2 threshold. The two kind of ARP scanning prevention can be start using at the same time, after port is banned, it can recover the state by configure the function of automatic recovery. After IP is banned, it can be automatic recovery when the rate of received arp packets is lower than level-2 threshold.

To improve the effect of the switch, users can configure trusted ports and IP, the ARP messages from which will not be checked by the switch. Thus the load of the switch can be effectively decreased.

3.8.2 ARP Scanning Prevention Configuration Task Sequence

1. Enable the ARP Scanning Prevention function.
2. Configure the threshold of the port-based and IP-based ARP Scanning Prevention
3. Configure trusted ports
4. Configure trusted IP
5. Configure automatic recovery time
6. Display relative information of debug information and ARP scanning
7. Configure the action after above level-2 threshold.

1. Enable the ARP Scanning Prevention function.

Command	Explanation
Global configuration mode	
anti-arpscan enable [ip|port] **no anti-arpscan enable [ip|port]**	Enable or disable the ARP Scanning Prevention function globally.

2. Configure the threshold of the port-based and IP-based ARP Scanning Prevention

Command	Explanation
Global configuration mode	
anti-arpscan port-based threshold *<threshold-value>* **no anti-arpscan port-based threshold**	Set the threshold of the port-based ARP Scanning Prevention.

anti-arpscan ip-based {level1|level2} threshold <*threshold-value*> **no anti-arpscan ip-based {level1|level2} threshold**	Set the threshold of the IP-based ARP Scanning Prevention.

3. Configure trusted ports

Command	Explanation
Port configuration mode	
anti-arpscan trust {port | supertrust-port | iptrust-port} **no anti-arpscan trust {port | supertrust-port | iptrust-port}**	Set the trust attributes of the ports.

4. Configure trusted IP

Command	Explanation
Global configuration mode	
anti-arpscan trust ip <*ip-address*> [<*netmask*>] **no anti-arpscan trust ip <*ip-address*> [<*netmask*>]**	Set the trust attributes of IP.

5. Configure automatic recovery time

Command	Explanation
Global configuration mode	
anti-arpscan recovery enable **no anti-arpscan recovery enable**	Enable or disable the automatic recovery function.
anti-arpscan recovery time <*seconds*> **no anti-arpscan recovery time**	Set automatic recovery time.

6. Display relative information of debug information and ARP scanning

anti-arpscan log enable **no anti-arpscan log enable**	Enable or disable the log function of ARP scanning prevention.
anti-arpscan trap enable [level1|level2] **no anti-arpscan trap enable [level1|level2]**	Enable or disable the SNMP Trap function of ARP scanning prevention.
anti-arpscan FFP max-num <num>	The maximum quantity of ARP scanning prevention function occupied FFP item.
show anti-arpscan [trust {ip | port | supertrust-port | iptrust-port} | prohibited {ip | port}]	Display the state of operation and configuration of ARP scanning prevention.
show anti-arpscan ip-based attack-list [history]	Display source information or history source information of ARP scanning attacks prevention.

show anti-arpscan ip-based running-config	Display the current configuration of arp scanning prevention.
clear anti-arpscan speed-limit< IP Address>	Flush ARP limited rate for specified host manually.
clear anti-arpscan ip-isolate<IP Address>	Flush IP business isolation for specified host manually.
clear anti-arpscan attack-list {ip < IP Address > | all }	Clear the ARP limit of the specific host or all the hosts manually.
clear anti-arpscan attack-history-list {ip < IP Address > | all }	Clear the history attacks source information of ARP scanning prevention manually.
Admin Mode	
debug anti-arpscan [port | ip] **no debug anti-arpscan [port | ip]**	Enable or disable the debug switch of ARP scanning prevention.

7. Configure the action after above level-2 threshold

Command	Explanation
Global configuration Mode	
anti-arpscan ip-based level2 action {isolate | discard-ARP}	After above level-2 threshold, users can configure ip business isolation and discard ARP packets.
anti-arpscan ip-based arp-to-cpu speed <pps> **no anti-arpscan ip-based arp-to-cpu speed**	Configure the rate of ARP send to CPU when level-1 threshold overrun.

3.8.3 ARP Scanning Prevention Typical Examples

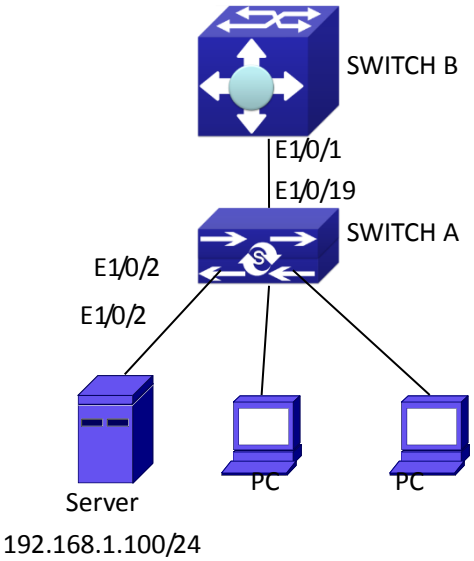


Figure 3-6 ARP scanning prevention typical configuration example

In the network topology above, port E1/0/1 of SWITCH B is connected to port E1/0/19 of

SWITCH A, the port E1/0/2 of SWITCH A is connected to file server (IP address is 192.168.1.100/24), and all the other ports of SWITCH A are connected to common PC. The following configuration can prevent ARP scanning effectively without affecting the normal operation of the system.
SWITCH A configuration task sequence:
SwitchA(config)#anti-arpscan enable
SwitchA(config)#anti-arpscan recovery time 3600
SwitchA(config)#anti-arpscan trust ip 192.168.1.100 255.255.255.0
SwitchA(config)#interface ethernet1/0/2
SwitchA (Config-If-Ethernet1/0/2)#anti-arpscan trust port
SwitchA (Config-If-Ethernet1/0/2)#exit
SwitchA(config)#interface ethernet1/0/19
SwitchA (Config-If-Ethernet1/0/19)#anti-arpscan trust supertrust-port
Switch A(Config-If-Ethernet1/0/19)#exit

SWITCHB configuration task sequence:
Switch B(config)# anti-arpscan enable
SwitchB(config)#interface ethernet1/0/1
SwitchB(Config-If-Ethernet1/0/1)#anti-arpscan trust port
SwitchB(Config-If-Ethernet1/0/1)exit

3.8.4 ARP Scanning Prevention Troubleshooting Help

☞ ARP scanning prevention is disabled by default. After enabling ARP scanning prevention, users can enable the debug switch, "**debug anti-arpscan**", to view debug information.

3.9 Prevent ARP Spoofing

3.9.1 Overview

3.9.1.1 ARP (Address Resolution Protocol)

Generally speaking, ARP (RFC-826) protocol is mainly responsible of mapping IP address to relevant 48-bit physical address, that is MAC address, for instance, IP address is 192.168.0.1, network card Mac address is 00-03-0F-FD-1D-2B. What the whole mapping process is that a host computer send broadcast data packet involving IP address information of destination host computer, ARP request, and then the destination host computer send a data packet involving its IP address and Mac address to the host, so two host computers can exchange data by MAC address.

3.9.1.2 ARP Spoofing

In terms of ARP Protocol design, to reduce redundant ARP data communication on networks, even though a host computer receives an ARP reply which is not requested by itself, it will also insert an entry to its ARP cache table, so it creates a possibility of “ARP spoofing”. If the hacker wants to snoop the communication between two host computers in the same network (even if are connected by the switches), it sends an ARP reply packet to two hosts separately, and make them misunderstand MAC address of the other side as the hacker host MAC address. In this way, the direct communication is actually communicated indirectly among the hacker host computer. The hackers not only obtain communication information they need, but also only need to modify some information in data packet and forward successfully. In this sniff way, the hacker host computer doesn’t need to configure intermix mode of network card, that is because the data packet between two communication sides are sent to hacker host computer on physical layer, which works as a relay.

3.9.1.3 How to prevent void ARP Spoofing

There are many sniff, monitor and attack behaviors based on ARP protocol in networks, and most of attack behaviors are based on ARP spoofing, so it is very important to prevent ARP spoofing. ARP spoofing accesses normal network environment by counterfeiting legal IP address firstly, and sends a great deal of counterfeited ARP application packets to switches, after switches learn these packets, they will cover previously corrected IP, mapping of MAC address, and then some corrected IP, MAC address mapping are modified to correspondence relationship configured by attack packets so that the switch makes mistake on transfer packets, and takes an effect on the whole network. Or the switches are made used of by vicious attackers, and they intercept and capture packets transferred by switches or attack other switches, host computers or network equipment.

What the essential method on preventing attack and spoofing switches based on ARP in networks is to disable switch automatic update function; the cheater can’t modify corrected MAC address in order to avoid wrong packets transfer and can’t obtain other information. At one time, it doesn’t interrupt the automatic learning function of ARP. Thus it prevents ARP spoofing and attack to a great extent.

3.9.2 Prevent ARP Spoofing configuration

The steps of preventing ARP spoofing configuration as below:

1. Disable ARP automatic update function
2. Disable ARP automatic learning function
3. Changing dynamic ARP to static ARP

1. Disable ARP automatic update function

Command	Explanation
Global Mode and Port Mode	

ip arp-security updateprotect **no ip arp-security updateprotect**	Disable and enable ARP automatic update function.

2. Disable ARP automatic learning function

Command	Explanation
Global mode and Interface Mode	
ip arp-security learnprotect **no ip arp-security learnprotect**	Disable and enable ARP automatic learning function.

3. Function on changing dynamic ARP to static ARP

Command	Explanation
Global Mode and Port Mode	
ip arp-security convert	Change dynamic ARP to static ARP.

3.9.3 Prevent ARP Spoofing Example

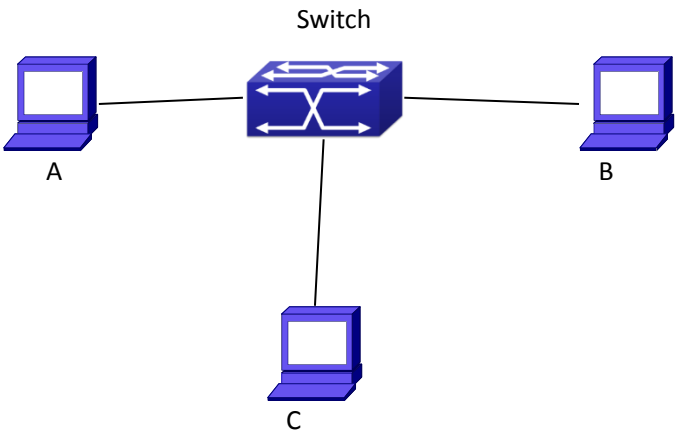


Figure 3-7

Equipment Explanation

Equipment	Configuration	Quality
switch	IP:192.168.2.4; IP:192.168.1.4; mac: 00-00-00-00-00-04	1
A	IP:192.168.2.1; mac: 00-00-00-00-00-01	1
B	IP:192.168.1.2; mac: 00-00-00-00-00-02	1
C	IP:192.168.2.3; mac: 00-00-00-00-00-03	some

There is a normal communication between B and C on above diagram. A wants switch to forward packets sent by B to itself, so need switch sends the packets transfer from B to A. firstly A sends ARP reply packet to switch, format is: 192.168.2.3, 00-00-00-00-00-01, mapping its MAC address to C's IP, so the switch changes IP address when it updates ARP list., then data packet of 192.168.2.3 is transferred to 00-00-00-00-00-01 address (A MAC address).

In further, a transfers its received packets to C by modifying source address and destination address, the mutual communicated data between B and C are received by A unconsciously.

Because the ARP list is update timely, another task for A is to continuously send ARP reply packet, and refreshes switch ARP list.

So it is very important to protect ARP list, configure to forbid ARP learning command in stable environment, and then change all dynamic ARP to static ARP, the learned ARP will not be refreshed, and protect for users.

```
Switch#config
Switch(config)#interface vlan 1
Switch(Config-If-Vlan1)#arp 192.168.2.1 00-00-00-00-00-01 interface eth 1/0/2
Switch(Config-If-Vlan1)#interface vlan 2
Switch(Config-If-Vlan2)#arp 192.168.1.2 00-00-00-00-00-02 interface eth 1/0/2
Switch(Config-If-Vlan2#interface vlan 3
Switch(Config-If-Vlan3)#arp 192.168.2.3 00-00-00-00-00-03 interface eth 1/0/2
Switch(Config-If-Vlan3)#exit
Switch(Config)#ip arp-security learnprotect
Switch(Config)#
Switch(config)#ip arp-security convert
```

If the environment changing, it enable to forbid ARP refresh, once it learns ARP property, it wont be refreshed by new ARP reply packet, and protect use data from sniffing.

```
Switch#config
Switch(config)#ip arp-security updateprotect
```

3.10 ARP GUARD

3.10.1 Introduction to ARP GUARD

There is serious security vulnerability in the design of ARP protocol, which is any network device, can send ARP messages to advertise the mapping relationship between IP address and MAC address. This provides a chance for ARP cheating. Attackers can send ARP REQUEST messages or ARP REPLY messages to advertise a wrong mapping relationship between IP address and MAC address, causing problems in network communication. The danger of ARP cheating has two forms: 1. PC4 sends an ARP message to advertise that the IP address of PC2 is mapped to the MAC address of PC4, which will cause all the IP messages to PC2 will be sent to PC4, thus PC4 will be able to monitor and capture the messages to PC2; 2. PC4 sends ARP messages to advertise that the IP address of PC2 is mapped to an illegal MAC address, which will prevent PC2 from receiving the messages to it. Particularly, if the attacker pretends to be the gateway and do ARP cheating, the whole network will be collapsed.

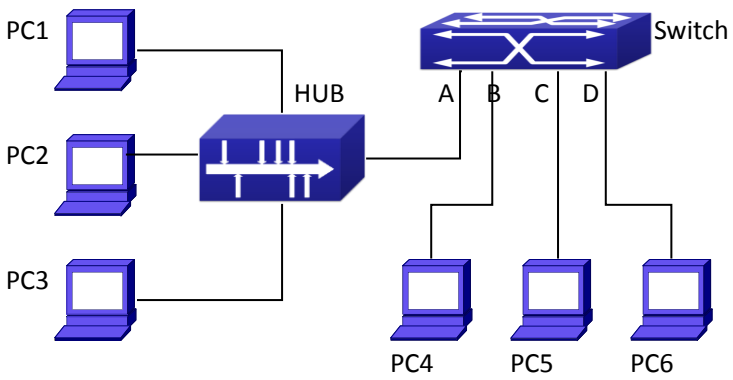


Figure 3-8 ARP GUARD schematic diagram

We utilize the filtering entries of the switch to protect the ARP entries of important network devices from being imitated by other devices. The basic theory of doing this is that utilizing the filtering entries of the switch to check all the ARP messages entering through the port, if the source address of the ARP message is protected, the messages will be directly dropped and will not be forwarded.

ARP GUARD function is usually used to protect the gateway from being attacked. If all the accessed PCs in the network should be protected from ARP cheating, then a large number of ARP GUARD address should be configured on the port, which will take up a big part of FFP entries in the chip, and as a result, might affect other applications. So this will be improper. It is recommended that adopting FREE RESOURCE related accessing scheme. Please refer to relative documents for details.

3.10.2 ARP GUARD Configuration Task List

1. Configure the protected IP address

Command	Explanation
Port configuration mode	
arp-guard ip <addr> **no arp-guard ip <addr>**	Configure/delete ARP GUARD address

3.11 ARP Local Proxy

3.11.1 Introduction to ARP Local Proxy function

In a real application environment, the switches in the aggregation layer are required to implement local ARP proxy function to avoid ARP cheating. This function will restrict the forwarding of ARP messages in the same vlan and thus direct the L3 forwarding of the data flow

through the switch.

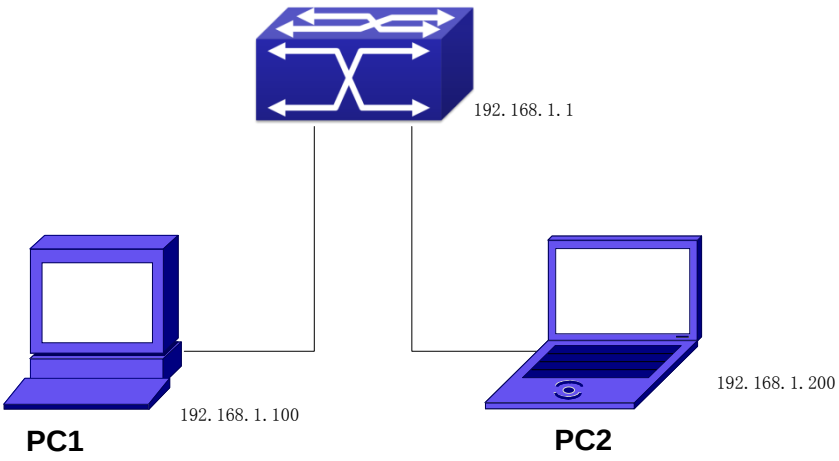


Figure 3-9

As shown in the figure above, PC1 wants to send an IP message to PC2, the overall procedure goes as follows (some non-arp details are ignored)

1. Since PC1 does not have the ARP of PC2, it sends and broadcasts ARP request.

2. Receiving the ARP message, the switch hardware will send the ARP request to CPU instead of forwarding this message via hardware, according to new ARP handling rules.

3. With local ARP proxy enabled, the switch will send ARP reply message to PC1 (to fill up its mac address)

4. After receiving the ARP reply, PC1 will create ARP, send an IP message, and set the destination MAC of the Ethernet head as the MAC of the switch.

5. After receiving the ip message, the switch will search the router table (to create router cache) and distribute hardware entries.

6. If the switch has the ARP of PC2, it will directly encapsulate the Ethernet head and send the message (the destination MAC is that of PC2)

7. If the switch does not have the ARP of PC2, it will request it and then send the ip message.

This function should cooperate with other security functions. When users configure local ARP proxy on an aggregation switch while configuring interface isolation function on the layer-2 switch connected to it, all ip flow will be forwarded on layer 3 via the aggregation switch. And due to the interface isolation, ARP messages will not be forwarded within the vlan, which means other PCs will not receive it.

3.11.2 ARP Local Proxy Function Configuration Task List

1. Enable/disable ARP local proxy function

1. Enable/disable ARP local proxy function

Command	Explanation
Interface vlan mode	

ip local proxy-arp **no ip local proxy-arp**	Enable or disable ARP local proxy function.

3.11.3 Typical Examples of ARP Local Proxy Function

As shown in the following figure, S1 is a medium/high-level layer-3 switch supporting ARP local proxy, S2 is layer-2 access switches supporting interface isolation.

Considering security, interface isolation function is enabled on S2. Thus all downlink ports of S2 is isolated from each other, making all ARP messages able to be forwarded through S1. If ARP local proxy is enabled on S1, then all interfaces on S1 isolate ARP while S1 serves as an ARP proxy. As a result, IP flow will be forwarded at layer 3 through S1 instead of S2.

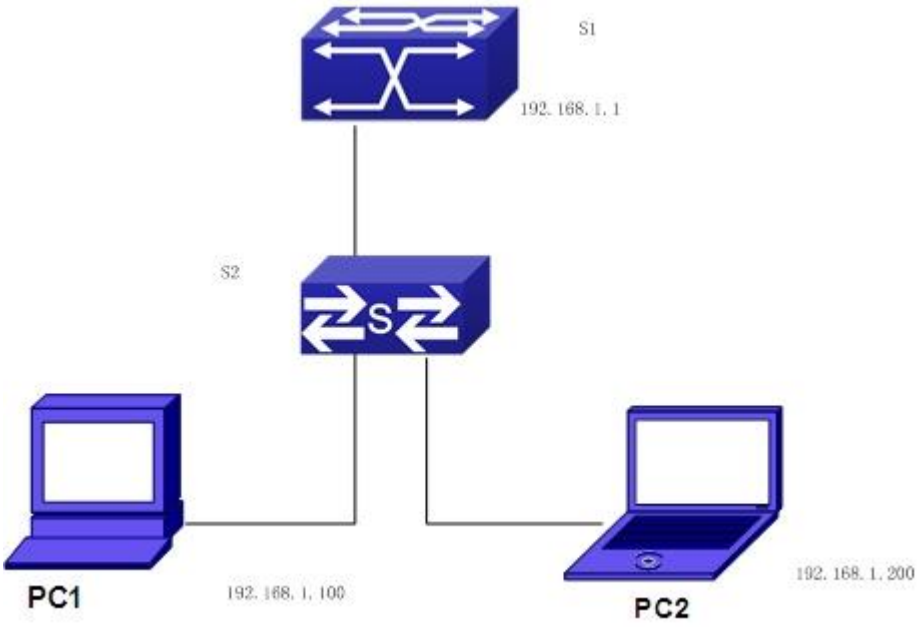


Figure 3-10

We can configure as follows:

Switch(config)#interface vlan 1

Switch(Config-if-Vlan1)#ip address 192.168.1.1 255.255.255.0

Switch(Config-if-Vlan1)#ip local proxy-arp

Switch(Config-if-Vlan1)#exit

3.11.4 ARP Local Proxy Function Troubleshooting

ARP local proxy function is disabled by default. Users can view the current configuration with display command. With correct configuration, by enabling debug of ARP, users can check whether the ARP proxy is normal and send proxy ARP messages.

In the process of operation, the system will show corresponding prompts if any operational error occurs.

3.12 Local arp

3.12.1 Introduction to Local ARP

When there are a large number of ARP packets passing through the device in a short time, the CPU load is too heavy, and the normal ARP can not be processed in time. When local-arp is enabled, only ARP request packets or reply packets whose destination IP is the interface IP of the device can trigger the device to learn ARP. Other passing ARP packets cannot trigger the device to learn ARP, and the packets will not be processed by the CPU, which can filter out the passing ARP packets and reduce the CPU load.

3.12.2 Local ARP Configuration Task List

1. enable local-arp function

Command	Explanation
Global configuration Mode	
local-arp enable **local-arp disable**	enable/disable Loca-arp function。

3.13 Gratuitous ARP

3.13.1 Introduction to Gratuitous ARP

Gratuitous ARP is a kind of ARP request that is sent by the host with its IP address as the destination of the ARP request.

The basic working mode for the switch is as below: The Layer 3 interfaces of the switch can be configured to advertise gratuitous ARP packets period or the switch can be configured to enable to send gratuitous ARP packets in all the interfaces globally.

The purpose of gratuitous ARP is as below:

1. To reduce the frequency that the host sends ARP request to the switch. The hosts in the network will periodically send ARP requests to the gateway to update the MAC address of the gateway. If the switch advertises gratuitous ARP requests, the host will not have to send these requests. This will reduce the frequency the hosts' sending ARP requests for the gateway's MAC address.
2. Gratuitous ARP is a method to prevent ARP cheating. The switch's advertising gratuitous ARP request will force the hosts to update its ARP table cache. Thus, forged ARP of gateway cannot function.

3.13.2 Gratuitous ARP Configuration Task List

1. Enable gratuitous ARP and configure the interval to send gratuitous ARP request
2. Display configurations about gratuitous ARP

1. Enable gratuitous ARP and configure the interval to send gratuitous ARP request.

Command	Explanation
Global Configuration Mode and Interface Configuration Mode.	
ip gratuitous-arp <*5-1200*> **no ip gratuitous-arp**	To enable gratuitous ARP and configure the interval to send gratuitous ARP request. The no command cancels the gratuitous ARP.

2. Display configurations about gratuitous ARP

Command	Explanation
Admin Mode and Configuration Mode	
show ip gratuitous-arp [interface vlan <*1-4094*>]	To display configurations about gratuitous ARP.

3.13.3 Gratuitous ARP Configuration Example

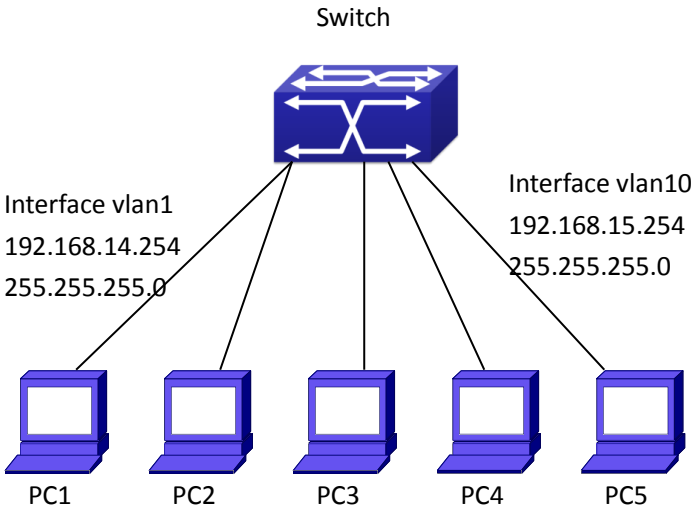


Figure 3-71 Gratuitous ARP Configuration Example

For the network topology shown in the figure above, interface VLAN10 whose IP address is 192.168.15.254 and network address mask is 255.255.255.0 in the switch system. Three PCs – PC3, PC4, PC5 are connected to the interface. The IP address of interface VLAN 1 is 192.168.14.254, its network address mask is 255.255.255.0. Two PCs – PC1 and PC2 are connected to this interface. Gratuitous ARP can be enabled through the following configuration:

1. Configure two interfaces to use gratuitous ARP at one time.

```
Switch(config)#ip gratuitous-arp 300
Switch(config)#exit
```

2. Configure gratuitous ARP specifically for only one interface at one time.

```
Switch(config)#interface vlan 10
Switch(Config-if-Vlan10)#ip gratuitous-arp 300
Switch(Config-if-Vlan10)#exit
Switch(config) #exit
```

3.13.4 Gratuitous ARP Troubleshooting

Gratuitous ARP is disabled by default. And when gratuitous ARP is enabled, the debugging information about ARP packets can be retrieved through the command debug ARP send.

If gratuitous ARP is enabled in global configuration mode, it can be disabled only in global configuration mode. If gratuitous ARP is configured in interface configuration mode, the configuration can only be disabled in interface configuration mode.

3.14 Keepalive Gateway

3.14.1 Introduction to Keepalive Gateway

Ethernet port is used to process backup or load balance, for the reason that it is a broadcast channel, it may not detect the change of physical signal and fails to get to down when the gateway is down. Keepalive Gateway is introduced to detect the connectivity to the higher-up gateway, in the case that a Ethernet port connect with a higher-up gateway to form a point-to-point network topology.

For example: router connects optical terminal device and the line is up all the time, While the line between moden and remote gateway is down, it is necessary to use a effective method to detect whether the remote gateway is reachable. At present, detect gateway connectivity by sending ARP request to gateway on time, if ARP resolution is failing, shutdown the interface, if ARP resolution is successful, keep the interface up.

Only layer 3 switch supports keepalive gateway function.

3.14.2 Keepalive Gateway Configuration Task List

1. Enable or disable keepalive gateway, configure the interval period that ARP request packet is sent and the retry-count after detection is failing
2. Show keepalive gateway and IPv4 running status of the interface

1. Enable or disable keepalive gateway, configure the interval period that ARP request packet is sent and the retry-count after detection is failing

Command	Explanation
Interface mode	
keepalive gateway *<ip-address>* [{*<interval-seconds>* | msec *<interval-millisecond >*} [retry-count]] **no keepalive gateway**	Enable keepalive gateway, configure IP address of gateway, the interval period that ARP request packet is sent, and the retry-count after detection is failing, the no command disables the function.

2. Show keepalive gateway and IPv4 running status of interface

Command	Explanation
Admin and configuration mode	
show keepalive gateway [interface-name]	Show keepalive running status of the specified interface, if there is no interface is specified, show keepalive running status of all interfaces.
show ip interface [interface-name]	Show IPv4 running status of the specified interface, if there is no interface is specified, show IPv4 running status of all interfaces.

3.14.3 Keepalive Gateway Example



Figure 3-82 keepalive gateway typical example

In above network topology, interface address of interface vlan10 is 1.1.1.1 255.255.255.0 for gateway A, interface address of interface vlan100 is 1.1.1.2 255.255.255.0 for gateway B, gateway B supports keepalive gateway function, the configuration in the following:

1. Adopt the default interval that ARP packet is sent and the retry-count after detection is failing (the default interval is 10s, the default retry-count is 5 times)

Switch(config)#interface vlan 100

Switch(config-if-vlan100)#keepalive gateway 1.1.1.1
Switch(config-if-vlan100)#exit
2. Configure the interval that ARP packet is sent and the retry-count after detection is failing manually.
Switch(config)#interface vlan 100
Switch(config-if-vlan100)#keepalive gateway 1.1.1.1 3 3
Switch(config-if-vlan100)#exit

Send ARP detection once 3 seconds to detect whether gateway A is reachable, after 3 times detection is failing, gateway A is considered to be unreachable.

3.14.4 Kepalive Gteway Troubleshooting

If there is any problem happens when using keepalive gateway function, please check whether the problem is caused by the following reasons:

- ☞ Make sure the device is layer 3 switch, layer 2 switch does not support keepalive gateway
- ☞ The detection method is used to point-to-point topology mode only
- ☞ Detect IPv4 accessibility by the method, so the detection result only affects IPv4 traffic, other traffic such as IPv6 is not affected
- ☞ Physical state of interface only controlled by physical signal
- ☞ Interface can't run IPv4 after determine gateway is not reachable, so all relative IPv4 routes are deleted and IPv4 route protocol can't establish the neighbor on the interface

3.15 DHCP

3.15.1 Introduction to DHCP

DHCP [RFC2131] is the acronym for Dynamic Host Configuration Protocol. It is a protocol that assigns IP address dynamically from the address pool as well as other network configuration parameters such as default gateway, DNS server, and default route and host image file position within the network. DHCP is the enhanced version of BOOTP. It is a mainstream technology that can not only provide boot information for diskless workstations, but can also release the administrators from manual recording of IP allocation and reduce user effort and cost on configuration. Another benefit of DHCP is it can partially ease the pressure on IP demands, when the user of an IP leaves the network that IP can be assigned to another user.

DHCP is a client-server protocol, the DHCP client requests the network address and configuration parameters from the DHCP server; the server provides the network address and configuration parameters for the clients; if DHCP server and clients are located in different subnets, DHCP relay is required for DHCP packets to be transferred between the DHCP client and DHCP server. The implementation of DHCP is shown below:

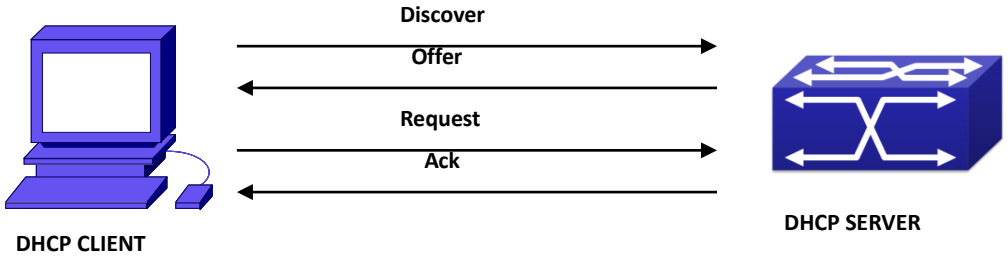


Figure 3-93 DHCP protocol interaction

Explanation:

1. DHCP client broadcasts DHCPDISCOVER packets in the local subnet.
2. On receiving the DHCPDISCOVER packet, DHCP server sends a DHCPOFFER packet along with IP address and other network parameters to the DHCP client.
3. DHCP client broadcast DHCPREQUEST packet with the information for the DHCP server it selected after selecting from the DHCPOFFER packets.
4. The DHCP server selected by the client sends a DHCPACK packet and the client gets an IP address and other network configuration parameters.

The above four steps finish a Dynamic host configuration assignment process. However, if the DHCP server and the DHCP client are not in the same network, the server will not receive the DHCP broadcast packets sent by the client, therefore no DHCP packets will be sent to the client by the server. In this case, a DHCP relay is required to forward such DHCP packets so that the DHCP packets exchange can be completed between the DHCP client and server.

Switch can act as both a DHCP server and a DHCP relay. DHCP server supports not only dynamic IP address assignment, but also manual IP address binding (i.e. specify a specific IP address to a specified MAC address or specified device ID over a long period. The differences and relations between dynamic IP address allocation and manual IP address binding are: 1) IP address obtained dynamically can be different every time; manually bound IP address will be the same all the time. 2) The lease period of IP address obtained dynamically is the same as the lease period of the address pool, and is limited; the lease of manually bound IP address is theoretically endless. 3) Dynamically allocated address cannot be bound manually. 4) Dynamic DHCP address pool can inherit the network configuration parameters of the dynamic DHCP address pool of the related segment.

3.15.2 DHCP Server Configuration

DHCP Sever Configuration Task List:

1. Enable/Disable DHCP service
2. Configure DHCP Address pool
   (1) Create/Delete DHCP Address pool
   (2) Configure DHCP address pool parameters
   (3) Configure manual DHCP address pool parameters

3. Enable logging for address conflicts

1. Enable/Disable DHCP service

Command	Explanation
Global Mode	
service dhcp **no service dhcp**	Enable DHCP server. The no command disables DHCP server.
ip dhcp disbale **no ip dhcp disable**	The port disables DHCP services, the no command enables DHCP services.

2. Configure DHCP Address pool

(1) Create/Delete DHCP Address pool

Command	Explanation
Global Mode	
ip dhcp pool <*name*> **no ip dhcp pool <*name*>**	Configure DHCP Address pool. The no operation cancels the DHCP Address pool.

(2) Configure DHCP address pool parameters

Command	Explanation
DHCP Address Pool Mode	
network-address <*network-number*> [mask | prefix-length] **no network-address**	Configure the address scope that can be allocated to the address pool. The no operation of this command cancels the allocation address pool.
default-router [<*address1*>[<*address2*>[…<*address8*>]]] **no default-router**	Configure default gateway for DHCP clients. The no operation cancels the default gateway.
dns-server [<*address1*>[<*address2*>[…<*address8*>]]] **no dns-server**	Configure DNS server for DHCP clients. The no command deletes DNS server configuration.
domain-name <*domain*> **no domain-name**	Configure Domain name for DHCP clients; the "**no domain-name**" command deletes the domain name.
netbios-name-server [<*address1*>[<*address2*>[…<*address8*>]]] **no netbios-name-server**	Configure the address for WINS server. The no operation cancels the address for server.
netbios-node-type {b-node|h-node|m-node|p-node|<*type-number*>} **no netbios-node-type**	Configure node type for DHCP clients. The no operation cancels the node type for DHCP clients.

bootfile *<filename>* **no bootfile**	Configure the file to be imported for DHCP clients on boot up. The no command cancels this operation.
next-server [*<address1>*[*<address2>*[…*<address8>*]]] **no next-server [*<address1>*[*<address2>*[…*<address8>*]]]**	Configure the address of the server hosting file for importing. The no command deletes the address of the server hosting file for importing.
option *<code>* {ascii *<string>* | hex *<hex>* | ipaddress *<ipaddress>*} **no option *<code>***	Configure the network parameter specified by the option code. The no command deletes the network parameter specified by the option code.
lease { days [hours][minutes] | infinite } **no lease**	Configure the lease period allocated to addresses in the address pool. The no command deletes the lease period allocated to addresses in the address pool.
max-lease-time {[*<days>*] [*<hours>*] [*<minutes>*] | infinite} **no max-lease-time**	Set the maximum lease time for the addresses in the address pool; the no command restores the default setting.
Global Mode	
ip dhcp excluded-address *<low-address>* [*<high-address>*] **no ip dhcp excluded-address *<low-address>* [*<high-address>*]**	Exclude the addresses in the address pool that are not for dynamic allocation.

(3) Configure manual DHCP address pool parameters

Command	Explanation
DHCP Address Pool Mode	
hardware-address *<hardware-address>* [{Ethernet | IEEE802|*<type-number>*}] **no hardware-address**	Specify/delete the hardware address when assigning address manually.
host *<address>* [*<mask>* / *<prefix-length>*] **no host**	Specify/delete the IP address to be assigned to the specified client when binding address manually.
client-identifier *<unique-identifier>* **no client-identifier**	Specify/delete the unique ID of the user when binding address manually.

3. Enable logging for address conflicts

Command	Explanation
Global Mode	
ip dhcp conflict logging **no ip dhcp conflict logging**	Enable/disable logging for DHCP address to detect address conflicts.
Admin Mode	
clear ip dhcp conflict *<address* / all >	Delete a single address conflict record or all conflict records.

3.15.3 DHCP Relay Configuration

When the DHCP client and server are in different segments, DHCP relay is required to transfer DHCP packets. Adding a DHCP relay makes it unnecessary to configure a DHCP server for each segment, one DHCP server can provide the network configuration parameter for clients from multiple segments, which is not only cost-effective but also management-effective.

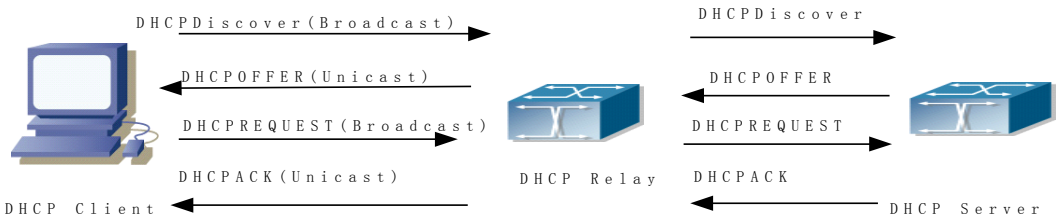


Figure 3-104 DHCP relay

As shown in the above figure, the DHCP client and the DHCP server are in different networks, the DHCP client performs the four DHCP steps as usual yet DHCP relay is added to the process.

1. The client broadcasts a DHCPDISCOVER packet, and DHCP relay inserts its own IP address to the relay agent field in the DHCPDISCOVER packet on receiving the packet, and forwards the packet to the specified DHCP server (for DHCP frame format, please refer to RFC2131).
2. On the receiving the DHCPDISCOVER packets forwarded by DHCP relay, the DHCP server sends the DHCPOFFER packet via DHCP relay to the DHCP client.
3. DHCP client chooses a DHCP server and broadcasts a DHCPREQUEST packet, DHCP relay forwards the packet to the DHCP server after processing.
4. On receiving DHCPREQUEST, the DHCP server responds with a DHCPACK packet via DHCP relay to the DHCP client.

DHCP Relay Configuration Task List:

1. Enable DHCP relay.
2. Configure DHCP relay to forward DHCP broadcast packet.

1. Enable DHCP relay.

Command	Explanation
Global Mode	
service dhcp **no service dhcp**	DHCP server and DHCP relay is enabled as the DHCP service is enabled.

2. Configure DHCP relay to forward DHCP broadcast packet.

Command	Explanation
Global Mode	
ip forward-protocol udp bootps **no ip forward-protocol udp bootps**	The UDP port 67 is used for DHCP broadcast packet forwarding.

Interface Configuration Mode	
ip helper-address *<ipaddress>* **no ip helper-address *<ipaddress>***	Set the destination IP address for DHCP relay forwarding; the "**no ip helper-address *<ipaddress>***"command cancels the setting.

3.15.4 DHCP Configuration Examples

Scenario 1:

Too save configuration efforts of network administrators and users, a company is using switch as a DHCP server. The Admin VLAN IP address is 10.16.1.2/16. The local area network for the company is divided into network A and B according to the office locations. The network configurations for location A and B are shown below.

PoolA(network 10.16.1.0)		PoolB(network 10.16.2.0)	
Device	IP address	Device	IP address
Default gateway	10.16.1.200 10.16.1.201	Default gateway	10.16.1.200 10.16.1.201
DNS server	10.16.1.202	DNS server	10.16.1.202
WINS server	10.16.1.209	WWW server	10.16.1.209
WINS node type	H-node		
Lease	3 days	Lease	1day

In location A, a machine with MAC address 00-03-22-23-dc-ab is assigned with a fixed IP address of 10.16.1.210 and named as "management".

Switch(config)#service dhcp
Switch(config)#interface vlan 1
Switch(Config-Vlan-1)#ip address 10.16.1.2 255.255.0.0
Switch(Config-Vlan-1)#exit
Switch(config)#ip dhcp pool A
Switch(dhcp-A-config)#network 10.16.1.0 24
Switch(dhcp-A-config)#lease 3
Switch(dhcp-A-config)#default-route 10.16.1.200 10.16.1.201
Switch(dhcp-A-config)#dns-server 10.16.1.202
Switch(dhcp-A-config)#netbios-name-server 10.16.1.209
Switch(dhcp-A-config)#netbios-node-type H-node
Switch(dhcp-A-config)#exit
Switch(config)#ip dhcp excluded-address 10.16.1.200 10.16.1.201
Switch(config)#ip dhcp pool B
Switch(dhcp-B-config)#network 10.16.2.0 24
Switch(dhcp-B-config)#lease 1
Switch(dhcp-B-config)#default-route 10.16.2.200 10.16.2.201
Switch(dhcp-B-config)#dns-server 10.16.2.202
Switch(dhcp-B-config)#option 72 ip 10.16.2.209
Switch(dhcp-config)#exit

Switch(config)#ip dhcp excluded-address 10.16.2.200 10.16.2.201
Switch(config)#ip dhcp pool A1
Switch(dhcp-A1-config)#host 10.16.1.210
Switch(dhcp-A1-config)#hardware-address 00-03-22-23-dc-ab
Switch(dhcp-A1-config)#exit

Usage Guide: When a DHCP/BOOTP client is connected to a VLAN1 port of the switch, the client can only get its address from 10.16.1.0/24 instead of 10.16.2.0/24. This is because the broadcast packet from the client will be requesting the IP address in the same segment of the VLAN interface after VLAN interface forwarding, and the VLAN interface IP address is 10.16.1.2/24, therefore the IP address assigned to the client will belong to 10.16.1.0/24.

If the DHCP/BOOTP client wants to have an address in 10.16.2.0/24, the gateway forwarding broadcast packets of the client must belong to 10.16.2.0/24. The connectivity between the client gateway and the switch must be ensured for the client to get an IP address from the 10.16.2.0/24 address pool.

Scenario 2:

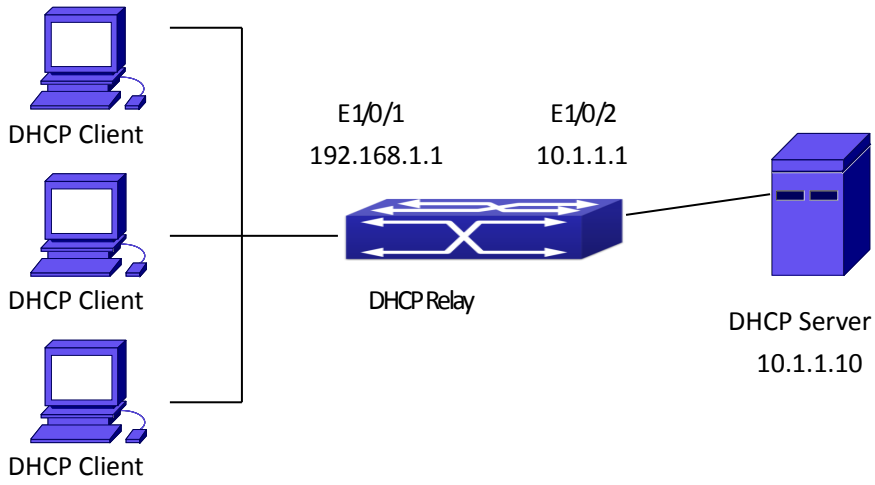


Figure 3-15 DHCP Relay Configuration

As shown in the above figure, route switch is configured as a DHCP relay. The DHCP server address is 10.1.1.10, the configuration steps is as follows:

Switch(config)#service dhcp
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 192.168.1.1 255.255.255.0
Switch(Config-if-Vlan1)#exit
Switch(config)#vlan 2
Switch(Config-Vlan-2)#exit
Switch(config)#interface Ethernet 1/0/2
Switch(Config-Erthernet1/0/2)#switchport access vlan 2
Switch(Config-Erthernet1/0/2)#exit
Switch(config)#interface vlan 2

```
Switch(Config-if-Vlan2)#ip address 10.1.1.1 255.255.255.0
Switch(Config-if-Vlan2)#exit
Switch(config)#ip forward-protocol udp bootps
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip help-address 10.1.1.10
Switch(Config-if-Vlan1)#exit
```

Note: It is recommended to use the combination of command **ip forward-protocol udp *<port>*** and **ip helper-address *<ipaddress>***. **ip help-address** can only be configured for ports on layer 3 and cannot be configured on layer 2 ports directly.

3.15.5 DHCP Troubleshooting

If the DHCP clients cannot obtain IP addresses and other network parameters, the following procedures can be followed when DHCP client hardware and cables have been verified ok.

- ☞ Verify the DHCP server is running, start the related DHCP server if not running. If the DHCP clients and servers are not in the same physical network, verify the router responsible for DHCP packet forwarding has DHCP relay function. If DHCP relay is not available for the intermediate router, it is recommended to replace the router or upgrade its software to one that has a DHCP relay function.
- ☞ In such case, DHCP server should be examined for an address pool that is in the same segment of the switch VLAN, such a pool should be added if not present, and (This does not indicate switch cannot assign IP address for different segments, see solution 2 for details.)
- ☞ In DHCP service, pools for dynamic IP allocation and manual binding are conflicting, i.e., if command "**network-address**" and "**host**" are run for a pool, only one of them will take effect; furthermore, in manual binding, only one IP-MAC binding can be configured in one pool. If multiple bindings are required, multiple manual pools can be created and IP-MAC bindings set for each pool. New configuration in the same pool overwrites the previous configuration.

3.16 DHCP option 82

3.16.1 Introduction to DHCP option 82

DHCP option 82 is the Relay Agent Information Option, its option code is 82. DHCP option 82 is aimed at strengthening the security of DHCP servers and improving the IP address configuration policy. The Relay Agent adds option 82 (including the client's physical access port, the access device ID and other information), to the DHCP request message from the client then forwards the message to DHCP server. When the DHCP server which supports the option 82 function receives the message, it will allocate an IP address and other configuration information for the client according to preconfigured policies and the option 82 information in the message.

At the same time, DHCP server can identify all the possible DHCP attack messages according to the information in option 82 and defend against them. DHCP Relay Agent will peel the option 82 from the reply messages it receives, and forward the reply message to the specified port of the network access device, according to the physical port information in the option. The application of DHCP option 82 is transparent for the client.

3.16.1.1 DHCP option 82 Message Structure

A DHCP message can have several option segments; option 82 is one of them. It has to be placed after other options but before option 255. The following is its format:

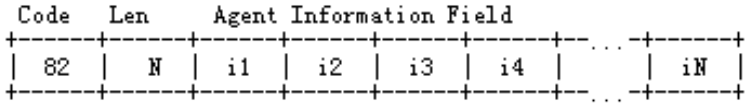


Figure 3-16

Code: represents the sequence number of the relay agent information option, the option 82 is called so because RFC3046 is defined as 82.

Len: the number of bytes in Agent Information Field, not including the two bytes in Code segment and Len segment.

Option 82 can have several sub-options, and need at least one sub-option. RFC3046 defines the following two sub-options, whose formats are showed as follows:

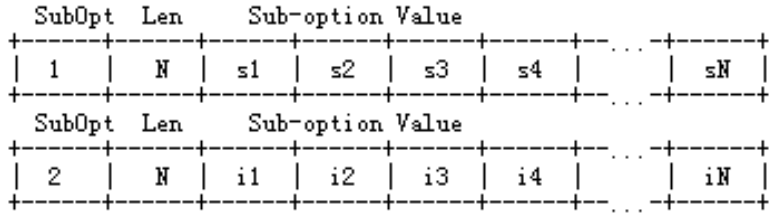


Figure 3-17

SubOpt: the sequence number of sub-option, the sequence number of Circuit ID sub-option is 1, the sequence number of Remote ID sub-option is 2.

Len: the number of bytes in Sub-option Value, not including the two bytes in SubOpt segment and Len segment.

3.16.1.2 option 82 Working Mechanism

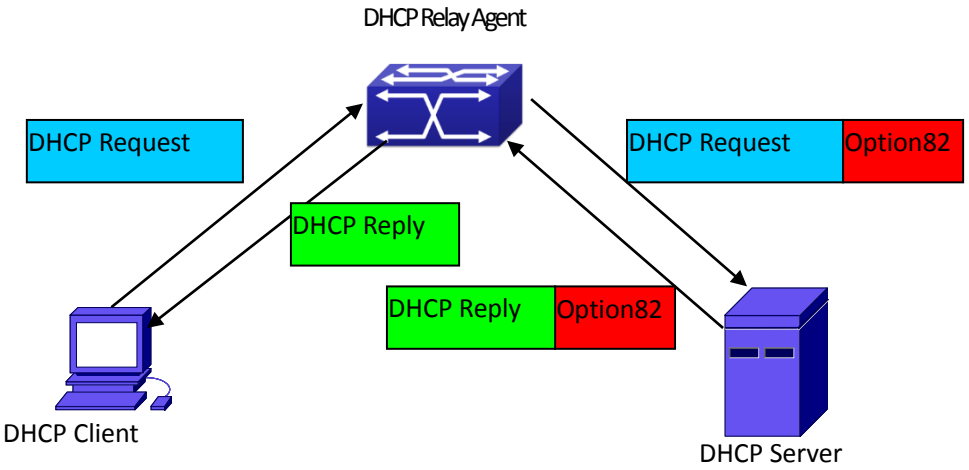


Figure 3-18 DHCP option 82 flow chart

If the DHCP Relay Agent supports option 82, the DHCP client should go through the following four steps to get its IP address from the DHCP server: discover, offer, select and acknowledge. The DHCP protocol follows the procedure below:

1）DHCP client sends a request broadcast message while initializing. This request message does not have option 82.

2）DHCP Relay Agent will add the option 82 to the end of the request message it receives, then relay and forward the message to the DHCP server. By default, the sub-option 1 of option 82 (Circuit ID) is the interface information of the switch connected to the DHCP client (VLAN name and physical port name), but the users can configure the Circuit ID as they wish. The sub-option 2 of option 82(Remote ID) is the MAC address of the DHCP relay device.

3）After receiving the DHCP request message, the DHCP server will allocate IP address and other information for the client according to the information and preconfigured policy in the option segment of the message. Then it will forward the reply message with DHCP configuration information and option 82 information to DHCP Relay Agent.

4）DHCP Relay Agent will peel the option 82 information from the replay message sent by DHCP server, and then forward the message with DHCP configuration information to the DHCP client.

3.16.2 DHCP option 82 Configuration Task List

1. Enabling the DHCP option 82 of the Relay Agent
2. Configure the DHCP option 82 attributes of the interface
3. Enable the DHCP option 82 of server
4. Configure DHCP option 82 default format of Relay Agent
5. Configure delimiter
6. Configure creation method of option82
7. Diagnose and maintain DHCP option 82

1. Enabling the DHCP option 82 of the Relay Agent.

Command	Explanation

Global mode	
ip dhcp relay information option **no ip dhcp relay information option**	Set this command to enable the option 82 function of the switch Relay Agent. The “no ip dhcp relay information option” is used to disable the option 82 function of the switch Relay Agent.

2. Configure the DHCP option 82 attributes of the interface

Command	Explanation
Interface configuration mode	
ip dhcp relay information policy {drop | keep | replace} **no ip dhcp relay information policy**	This command is used to set the retransmitting policy of the system for the received DHCP request message which contains option 82. The drop mode means that if the message has option82, then the system will drop it without processing; keep mode means that the system will keep the original option 82 segment in the message, and forward it to the server to process; replace mode means that the system will replace the option 82 segment in the existing message with its own option 82, and forward the message to the server to process. The “no ip dhcp relay information policy” will set the retransmitting policy of the option 82 DCHP message as “replace”.
ip dhcp relay information option subscriber-id {standard | *<circuit-id>*} **no ip dhcp relay information option subscriber-id**	This command is used to set the format of option 82 sub-option1(Circuit ID option) added to the DHCP request messages from interface, standard means the standard VLAN name and physical port name format, like”Vlan2+Ethernet1/0/12”,<circuit-id> is the circuit-id contents of option 82 specified by users, which is a string no longer than 64characters. The” **no ip dhcp relay information option subscriber-id**” command will set the format of added option 82 sub-option1 (Circuit ID option) as standard format.
Global Mode	

ip dhcp relay information option remote-id {standard | *<remote-id>*} **no ip dhcp relay information option remote-id**	Set the suboption2 (remote ID option) content of option 82 added by DHCP request packets (They are received by the interface). The no command sets the additive suboption2 (remote ID option) format of option 82 as standard.

3. Enable the DHCP option 82 of server.

Command	Explanation
Global mode	
ip dhcp server relay information enable **no ip dhcp server relay information enable**	This command is used to enable the switch DHCP server to identify option82. The "**no ip dhcp server relay information enable**" command will make the server ignore the option 82.

4. Configure DHCP option 82 default format of Relay Agent

Command	Explanation
Global mode	
ip dhcp relay information option subscriber-id format {hex | acsii | vs-hp}	Set subscriber-id format of Relay Agent option82.
ip dhcp relay information option remote-id format {default | vs-hp}	Set remote-id format of Relay Agent option82.

5. Configure delimiter

Command	Explanation
Global mode	
ip dhcp relay information option delimiter [colon | dot | slash | space] **no ip dhcp relay information option delimiter**	Set the delimiter of each parameter for suboption of option82 in global mode, no command restores the delimiter as slash.

6. Configure creation method of option82

Command	Explanation
Global mode	
ip dhcp relay information option self-defined remote-id {hostname | mac | string WORD} **no ip dhcp relay information option self-defined remote-id**	Set creation method for option82, users can define the parameters of remote-id suboption by themselves
ip dhcp relay information option self-defined remote-id format [ascii | hex]	Set self-defined format of remote-id for relay option82.

ip dhcp relay information option self-defined subscriber-id {vlan | port | id (switch-id (mac | hostname)| remote-mac)| string WORD } **no ip dhcp relay information option self-defined subscriber-id**	Set creation method for option82, users can define the parameters of circute-id suboption by themselves
ip dhcp relay information option self-defined subscriber-id format [ascii | hex]	Set self-defined format of circuit-id for relay option82.

7. Diagnose and maintain DHCP option 82

Command	Explanation
Admin mode	
show ip dhcp relay information option	This command will display the state information of the DHCP option 82 in the system, including option82 enabling switch, the interface retransmitting policy, the circuit ID mode and the DHCP server option82 enabling switch.
debug ip dhcp relay packet	This command is used to display the information of data packets processing in DHCP Relay Agent, including the “add” and “peel” action of option 82.

3.16.3 DHCP option 82 Application Examples

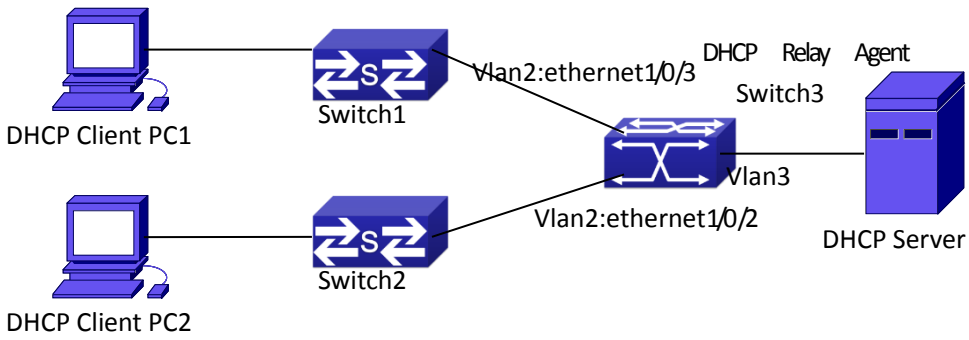


Figure 3-129 A DHCP option 82 typical application example

In the above example, layer 2 switches Switch1 and Switch2 are both connected to layer 3 switch Switch3, Switch 3 will transmit the request message from DHCP client to DHCP serer as DHCP Relay Agent. It will also transmit the reply message from the server to DHCP client to finish the DHCP protocol procedure. If the DHCP option 82 is disabled, DHCP server cannot distinguish that whether the DHCP client is from the network connected to Switch1 or Switch2. So, all the PC terminals connected to Switch1 and Switch2 will get addresses from the public address pool of the DHCP server. After the DHCP option 82 function is enabled, since the Switch3 appends the port information of accessing Switch3 to the request message from the client, the server can tell

that whether the client is from the network of Swich1 or Swich2, and thus can allocate separate address spaces for the two networks, to simplify the management of networks.

The following is the configuration of Switch3(MAC address is 00:03:0f:02:33:01):

```
Switch3(Config)#service dhcp
Switch3(Config)#ip dhcp relay information option
Switch3(Config)#ip forward-protocol udp bootps
Switch3(Config)#interface vlan 3
Switch3(Config-if-vlan3)#ip address 192.168.10.222 255.255.255.0
Switch3(Config-if-vlan2)#ip address 192.168.102.2 255.255.255.0
Switch3(Config-if-vlan2)#ip helper 192.168.10.88
```

Linux ISC DHCP Server supports option 82, its configuration file /etc/dhcpd.con is

```
ddns-update-style interim;
ignore client-updates;

class "Switch3Vlan2Class1" {
match if option agent.circuit-id = "Vlan2+Ethernet1/0/2" and option agent.remote-id=00:03:0f:02:33:01;
}

class "Switch3Vlan2Class2" {
match if option agent.circuit-id = "Vlan2+Ethernet1/0/3" and option agent.remote-id=00:03:0f:02:33:01;
}

subnet 192.168.102.0 netmask 255.255.255.0 {
option routers 192.168.102.2;
option subnet-mask 255.255.255.0;
option domain-name "example.com.cn";
option domain-name-servers 192.168.10.3;
authoritative;

pool {
range 192.168.102.21 192.168.102.50;
default-lease-time 86400; #24 Hours
max-lease-time 172800; #48 Hours
allow members of "Switch3Vlan2Class1";
}
pool {
range 192.168.102.51 192.168.102.80;
default-lease-time 43200; #12 Hours
max-lease-time 86400; #24 Hours
allow members of "Switch3Vlan2Class2";
```

}
}

Now, the DHCP server will allocate addresses for the network nodes from Switch1 which are relayed by Switch3 within the range of 192.168.102.21 ~ 192.168.102.50, and allocate addresses for the network nodes from Switch1 within the range of 192.168.102.51～192.168.102.80.

3.16.4 DHCP option 82 Troubleshooting

- DHCP option 82 is implemented as a sub-function module of DHCP Relay Agent. Before using it, users should make sure that the DHCP Relay Agent is configured correctly.
- DHCP option 82 needs the DHCP Relay Agent and the DHCP server cooperate to finish the task of allocating IP addresses. The DHCP server should set allocating policy correctly depending on the network topology of the DHCP Relay Agent, or, even the Relay Agent can operate normally, the allocation of addresses will fail. When there is more than one kind of Relay Agent, please pay attention to the retransmitting policy of the interface DHCP request messages.
- To implement the option 82 function of DHCP Relay Agent, the “debug dhcp relay packet” command can be used during the operating procedure, including adding the contents of option 82, the retransmitting policy adopted, the option 82 contents of the server peeled by the Relay Agent and etc., such information can help users to do troubleshooting.
- To implement the option 82 function of DHCP server, the “debug ip dhcp server packet” command can be used during the operating procedure to display the procedure of data packets processing of the server, including displaying the identified option 82 information of the request message and the option 82 information returned by the reply message.

3.17 DHCP Snooping

3.17.1 Introduction to DHCP Snooping

DHCP Snooping means that the switch monitors the IP-getting process of DHCP CLIENT via DHCP protocol. It prevents DHCP attacks and illegal DHCP SERVER by setting trust ports and untrust ports. And the DHCP messages from trust ports can be forwarded without being verified. In typical settings, trust ports are used to connect DHCP SERVER or DHCP RELAY Proxy, and untrust ports are used to connect DHCP CLINET. The switch will forward the DCHP request messages from untrust ports, but not DHCP reply ones. If any DHCP reply messages is received from a untrust port, besides giving an alarm, the switch will also implement designated actions

on the port according to settings, such as “shutdown”, or distributing a “blackhole”. If DHCP Snooping binding is enabled, the switch will save binding information (including its MAC address, IP address, IP lease, VLAN number and port number) of each DHCP CLINET on untrust ports in DHCP snooping binding table With such information, DHCP Snooping can combine modules like dot1x and ARP, or implement user-access-control independently.

Defense against Fake DHCP Server: once the switch intercepts the DHCP Server reply packets （including DHCPOFFER, DHCPACK, and DHCPNAK）, it will alarm and respond according to the situation（shutdown the port or send Black hole）。

Defense against DHCP over load attacks: To avoid too many DHCP messages attacking CPU, users should limit the DHCP speed of receiving packets on trusted and non-trusted ports.

Record the binding data of DHCP: DHCP SNOOPING will record the binding data allocated by DHCP SERVER while forwarding DHCP messages, it can also upload the binding data to the specified server to backup it. The binding data is mainly used to configure the dynamic users of dot1x user based ports. Please refer to the chapter called“dot1x configuration” to find more about the usage of dot1x use-based mode.

Add binding ARP: DHCP SNOOPING can add static binding ARP according to the binding data after capturing binding data, thus to avoid ARP cheating.

Add trusted users: DHCP SNOOPING can add trusted user list entries according to the parameters in binding data after capturing binding data; thus these users can access all resources without DOT1X authentication.

Automatic Recovery: A while after the switch shut down the port or send blockhole, it should automatically recover the communication of the port or source MAC and send information to Log Server via syslog.

LOG Function: When the switch discovers abnormal received packets or automatically recovers, it should send syslog information to Log Server.

The Encryption of Private Messages: The communication between the switch and the inner network security management system TrustView uses private messages. And the users can encrypt those messages of version 2.

Add authentication option82 Function: It is used **with dot1x** dhcpoption82 authentication mode. Different option 82 will be added in DHCP messages according to user’s authentication status.

3.17.2 DHCP Snooping Configuration Task Sequence

1. Enable DHCP Snooping
2. Enable DHCP Snooping binding function
3. Enable DHCP Snooping binding ARP function
4. Enable DHCP Snooping option82 function
5. Set the private packet version
6. Set DES encrypted key for private packets
7. Set helper server address
8. Set trusted ports
9. Enable DHCP Snooping binding DOT1X function

10. Enable DHCP Snooping binding USER function
11. Adding static list entries function
12. Set defense actions
13. Set rate limitation of DHCP messages
14. Enable the debug switch
15. Configure DHCP Snooping option 82 attributes

1. Enable DHCP Snooping

Command	Explanation
Globe mode	
ip dhcp snooping enable **no ip dhcp snooping enable**	Enable or disable the DHCP snooping function.

2. Enable DHCP Snooping binding

Command	Explanation
Globe mode	
ip dhcp snooping binding enable **no ip dhcp snooping binding enable**	Enable or disable the DHCP snooping binding function.

3. Enable DHCP Snooping binding ARP function

Command	Explanation
Globe mode	
ip dhcp snooping binding arp **no ip dhcp snooping binding arp**	Enable or disable the dhcp snooping binding ARP function.

4. Enable DHCP Snooping option82 function

Command	Explanation
Globe mode	
ip dhcp snooping information enable **no ip dhcp snooping information enable**	Enable/disable DHCP Snooping option 82 function.

5. Set the private packet version

Command	Explanation
Globe mode	
ip user private packet version two **no ip user private packet version two**	To configure/delete the private packet version.

6. Set DES encrypted key for private packets

Command	Explanation
Globe mode	
enable trustview key 0/7 <*password*> **no enable trustview key**	To configure/delete DES encrypted key for private packets.

7. Set helper server address

Command	Explanation
Globe mode	
ip user helper-address A.B.C.D [port <udpport>] source <ipAddr> (secondary|) **no ip user helper-address (secondary|)**	Set or delete helper server address.

8. Set trusted ports

Command	Explanation
Port mode	
ip dhcp snooping trust **no ip dhcp snooping trust**	Set or delete the DHCP snooping trust attributes of ports.

9. Enable DHCP SNOOPING binding DOT1X function

Command	Explanation
Port mode	
ip dhcp snooping binding dot1x **no ip dhcp snooping binding dot1x**	Enable or disable the DHCP snooping binding dot1x function.

10. Enable or disable the DHCP SNOOPING binding USER function

Command	Explanation
Port mode	
ip dhcp snooping binding user-control **no ip dhcp snooping binding user-control**	Enable or disable the DHCP snooping binding user function.

11. Add static binding information

Command	Explanation
Globe mode	
ip dhcp snooping binding user *<mac>* address *<ipAddr>* vlan ***<vid>*** **interface (ethernet|) *<ifname>*** **no ip dhcp snooping binding user *<mac>* interface (ethernet|) *<ifname>***	Add/delete DHCP snooping static binding list entries.

12. Set defense actions

Command	Explanation
Port mode	

ip dhcp snooping action {shutdown|blackhole} [recovery <*second*>] **no ip dhcp snooping action**	Set or delete the DHCP snooping automatic defense actions of ports.

13. Set rate limitation of data transmission

Command	Explanation
Globe mode	
ip dhcp snooping limit-rate <*pps*> **no ip dhcp snooping limit-rate**	Set rate limitation of the transmission of DHCP snooping messages.

14. Enable the debug switch

Command	Explanation
Admin mode	
debug ip dhcp snooping packet **debug ip dhcp snooping event** **debug ip dhcp snooping update** **debug ip dhcp snooping binding**	Please refer to the chapter on system troubleshooting.

15. Configure DHCP Snooping option 82 attributes

Command	Explanation
Globe mode	
ip dhcp snooping information option subscriber-id format {hex | acsii | vs-hp}	This command is used to set subscriber-id format of DHCP snooping option82.
ip dhcp snooping information option remote-id {standard | <*remote-id*>} **no ip dhcp snooping information option remote-id**	Set the suboption2 (remote ID option) content of option 82 added by DHCP request packets (they are received by the port). The no command sets the additive suboption2 (remote ID option) format of option 82 as standard.
ip dhcp snooping information option delimiter [colon | dot | slash | space] **no ip dhcp snooping information option delimiter**	Set the delimiter of each parameter for suboption of option82 in global mode, no command restores the delimiter as slash.
ip dhcp snooping information option self-defined remote-id {hostname | mac | string WORD} **no ip dhcp snooping information option self-defined remote-id**	Set creation method for option82, users can define the parameters of remote-id suboption by themselves.
ip dhcp snooping information option self-defined remote-id format [ascii | hex]	Set self-defined format of remote-id for snooping option82.

ip dhcp snooping information option self-defined subscriber-id {vlan | port | id (switch-id (mac | hostname)| remote-mac) | string WORD} **no ip dhcp snooping information option type self-defined subscriber-id**	Set creation method for option82, users can define the parameters of circute-id suboption by themselves.
ip dhcp snooping information option self-defined subscriber-id format [ascii | hex]	Set self-defined format of circuit-id for snooping option82.
Port mode	
ip dhcp snooping information option subscriber-id {standard | <*circuit-id*>} **no ip dhcp snooping information option subscriber-id**	Set the suboption1 (circuit ID option) content of option 82 added by DHCP request packets (they are received by the port). The no command sets the additive suboption1 (circuit ID option) format of option 82 as standard.

Command	Explanation
Globe mode	
ip dhcp snooping information option allow-untrusted (replace|) **no ip dhcp snooping information option allow-untrusted (replace|)**	This command is used to set that allow untrusted ports of DHCP snooping to receive DHCP packets with option82 option. When the "replace" is setting, the potion82 option is allowed to replace. When disabling this command, all untrusted ports will drop DHCP packets with option82 option.

3.17.3 DHCP Snooping Typical Application

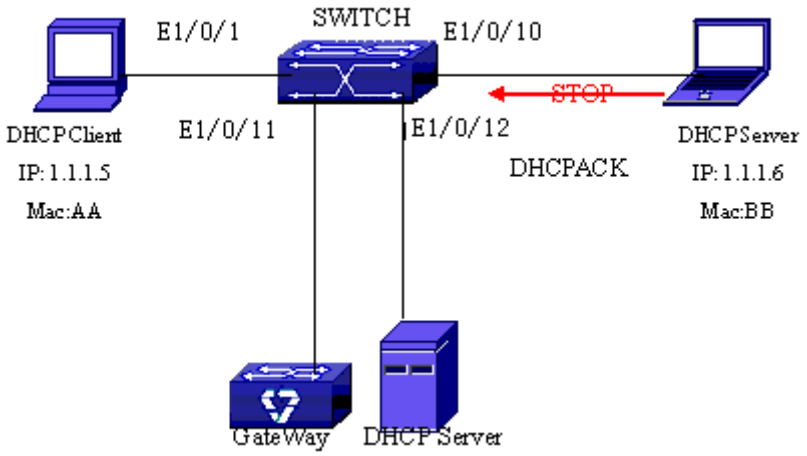


Figure 3-20 Sketch Map of TRUNK

As showed in the above chart, Mac-AA device is the normal user, connected to the

non-trusted port 1/0/1 of the switch. It operates via DHCP Client, IP 1.1.1.5; DHCP Server and GateWay are connected to the trusted ports 1/0/11 and 1/0/12 of the switch; the malicious user Mac-BB is connected to the non-trusted port 1/0/10, trying to fake a DHCP Server（by sending DHCPACK）. Setting DHCP Snooping on the switch will effectively detect and block this kind of network attack.

Configuration sequence is:

```
switch#
switch#config
switch(config)#ip dhcp snooping enable
switch(config)#interface ethernet 1/0/11
switch(Config-Ethernet1/0/11)#ip dhcp snooping trust
switch(Config-Ethernet1/0/11)#exit
switch(config)#interface ethernet 1/0/12
switch(Config-Ethernet1/0/12)#ip dhcp snooping trust
switch(Config-Ethernet1/0/12)#exit
switch(config)#interface ethernet 1/0/1-10
switch(Config-Port-Range)#ip dhcp snooping action shutdown
switch(Config-Port-Range)#
```

3.17.4 DHCP Snooping Troubleshooting Help

3.17.4.1 Monitor and Debug Information

The “debug ip dhcp snooping” command can be used to monitor the debug information.

3.17.4.2 DHCP Snooping Troubleshooting Help

If there is any problem happens when using DHCP Snooping function, please check if the problem is caused by the following reasons:

- ☞ Check that whether the global DHCP Snooping is enabled;
- ☞ If the port does not react to invalid DHCP Server packets, please check that whether the port is set as a non-trusted port of DHCP Snooping.

3.18 DHCP option 60 and option 43

3.18.1 Introduction to DHCP option 60 and option 43

DHCP server analyzes DHCP packets from DHCP client. If packets with option 60, it will decide whether option 43 is returned to DHCP client according to option 60 of packets and

configuration of option 60 and option 43 in DHCP server address pool.

Configure the corresponding option 60 and option 43 in DHCP server address pool:

1. Address pool configured option 60 and option 43 at the same time. The received DHCP packet with option 60 from DHCP client, if it matches with option 60 of DHCP server address pool, DHCP client will receive the option 43 configured in the address pool, or else do not return option 43 to DHCP client.
2. Address pool only configured option 43, it will match with any option 60. If the received DHCP packet with option 60 from DHCP client, DHCP client will receive the option 43 configured in the address pool.
3. Address pool only configured option 60, it will not return option 43 to DHCP client.

3.18.2 DHCP option 60 and option 43 Configuration Task List

1. Basic DHCP option 60 and option 43 configuration

Command	Explanation
Address pool configuration mode	
option 60 ascii LINE	Configure option 60 character string with ascii format in ip dhcp pool mode.
option 43 ascii LINE	Configure option 43 character string with ascii format in ip dhcp pool mode.
option 60 hex WORD	Configure option 60 character string with hex format in ip dhcp pool mode.
option 43 hex WORD	Configure option 43 character string with hex format in ip dhcp pool mode.
option 60 ip A.B.C.D	Configure option 60 character string with IP format in ip dhcp pool mode.
option 43 ip A.B.C.D	Configure option 43 character string with IP format in ip dhcp pool mode.
no option 60	Delete the configured option 60 in the address pool mode.
no option 43	Delete the configured option 43 in the address pool mode.

3.18.3 DHCPv6 option 60 and option 43 Example

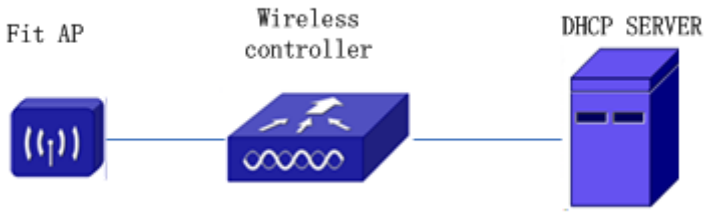


Figure 3-21 Typical DHCP option 60 and option 43 topology

Fit AP obtains IP address and option 43 attribute by DHCP server to send unicast discovery request for wireless controller. DHCP server configures option 60 matched with the option 60 of fit ap to return option 43 attribute to FTP AP. The wireless controller addresses of DHCP option 43 are 192.168.10.5 and 192.168.10.6.

Configuration procedure:

Configure DHCP server

switch (config)#ip dhcp pool a

switch (dhcp-a-config)#option 60 ascii AP1000

switch (dhcp-a-config)#option 43 hex 0104C0A80A050104C0A80A06

3.18.4 DHCP option 60 and option 43 Troubleshooting

If problems occur when configuring DHCP option 60 and option 43, please check whether the problem is caused by the following reasons:

- Check whether service dhcp function is enabled
- If the address pool configured option 60, check whether it matches with the option 60 of the packets.

Chapter 4 Routing Protocol Related Configuration

4.1 Routing Protocol Overview

To communicate with a remote host over the Internet, a host must choose a proper route via a set of routers or Layer3 switches.

Both routers and layer3 switches calculate the route using CPU, the difference is that layer3 switch adds the calculated route to the switch chip and forward by the chip at wire speed, while the router always store the calculated route in the route table or route buffer, and data forwarding is performed by the CPU. For this reason, although both routers and switches can perform route selection, layer3 switches have great advantage over routers in data forwarding. The following describes basic principle and methods used in layer3 switch route selection.

In route selection, the responsibility of each layer3 switch is to select a proper midway route according to the destination of the packet received; and send the packet to the next layer3 switch until the last layer3 switch in the route send the packet to the destination host. A route is the path selected by each layer3 switch to pass the packet to the next layer3 switch. Route can be grouped into direct route, static route and dynamic route.

Direct route refer to the path directly connects to the layer3 switch, and can be obtained with no calculation.

Static route is the manually specified path to a network or a host; static route cannot be changed freely. The advantage of static route is simple and consistent, and it can limit illegal route modification, and is convenient for load balance and route backup. However, as this is set manually, it is not suitable for mid- or large-scale networks for the route in such conditions are too huge and complex.

Dynamic route is the path to a network or a host calculated by the layer3 switch according to the routing protocols enabled. If the next hop layer3 switch in the path is not reachable, layer3 switch will automatically discard the path to that next hop layer3 switch and choose the path through other layer3 switches.

There are two dynamic routing protocols: Interior Gateway Protocol (IGP) and Exterior Gateway protocol (EGP). IGP is the protocol used to calculate the route to a destination inside an autonomous system. IGP supported by switch include RIP and OSPF, RIP and OSRF can be configured according to the requirement. Switch supports running several IGP dynamic routing protocols at the same time. Or, other dynamic routing protocols and static route can be introduced to a dynamic routing protocol, so that multiple routing protocols can be associated.

EGP is used to exchange routing information among different autonomous systems, such as BGP protocol. EGP supported by switch include BGP-4, BGP-4+.

4.1.1 Routing Table

As mentioned before, layer3 switch is mainly used to establish the route from the current layer3 switch to a network or a host, and to forward packets according to the route. Each layer3 switch has its own route table containing all routes used by that switch. Each route entry in the route table specifies the physical port should be used for forwarding packet to reach a destination host or the next hop layer3 switch to the host.

The route table mainly consists of the following:

- Destination address: used to identify the destination address or destination network of an IP packet.
- Network mask: used together with destination address to identify the destination host or the network the layer3 switch resides. Network mask consists of several consecutive binary 1's, and usually in the format of dotted decimal (an address consists of 1 to 4 255's.) When "AND" the destination address with network mask, we can get the network address for the destination host or the network the layer3 switch resides. For example, the network address of a host or the segment the layer3 switch resides with a destination address of 200.1.1.1 and mask 255.255.255.0 is 200.1.1.0.
- Output interface: specify the interface of layer3 switch to forward IP packets.
- IP address of the next layer3 switch (next hop): specify the next layer3 switch the IP packet will pass.
- Route entry priority: There may be several different next hop routes leading to the same destination. Those routes may be discovered by different dynamic routing protocols or static routes manually configured. The entry with the highest priority (smallest value) becomes the current best route. The user can configure several routes of different priority to the same destination; layer3 switch will choose one route for IP packet forwarding according to the priority order.

To prevent too large route table, a default route can be set. Once route table look up fails, the default route will be chosen for forwarding packets.

The table below describes the routing protocols supported by switch and the default route look up priority value.

Routing Protocols or route type	Default priority value
Direct route	0
OSPF	110
Static route	1
RIP	120
OSPF ASE	150
IBGP	200
EBGP	20
Unknown route	255

4.1.2 IP Routing Policy

4.1.2.1 Introduction to Routing Policy

Some policies have to be applied when the router publishing and receiving routing messages so to filter routing messages, such as only receiving or publishing routing messages meets the specified conditions. A routing protocol maybe need redistribute other routing messages found by other protocols such as OSPF so to increase its own routing knowledge; when the router redistributing routing messages from other routing protocols there may be only part of the qualified routing messages is needed, and some properties may have to be configured to suit this protocol.

To achieve routing policy, first we have to define the characteristics of the routing messages to be applied with routing policies, namely define a group matching rules. We can configure by different properties in the routing messages such as destination address, the router address publishing the routing messages. The matching rules can be previously configured to be applied in the routing publishing, receiving and distributing policies.

Five filters are provided in switch: route-map, acl, as-path, community-list and ip-prefix for use. We will introduce each filter in following sections:

1. route-map

For matching certain properties of the specified routing information and setting some routing propertities when the conditions are fulfilled.

Route-map is for controlling and changing the routing messages while also controlling the redistribution among routes. A route-map consists of a series of match and set commands in which the match command specifies the conditions required matching, and the set command specifies the actions to be taken when matches. The route-map is also for controlling route publishing among different route process. It can also used on policy routing which select different routes for the messages other than the shortest route.

A group matches and set clauses make up a node. A route-map may consist of several nodes each of which is a unit for matching test. We match among nodes with by sequence-number. Match clauses define matching rules. The matching objects are some properties of routing messages. Different match clause in the same node is “and” relation logically, which means the matching test of a node, will not be passed until conditions in its entire match clause are matched. Set clause specifies actions, namely configure some properties of routing messages after the matching test is passed.

Different nodes in a route-map is an “or” relation logically. The system checks each node of the route-map in turn and once certain node test is passed the route-map test will be passed without taking the next node test.

2. access control list(acl)

ACL (Access Control Lists) is a data packet filter mechanism in the switch. The switch controls the network access and secure the network service by permitting or denying certain data packet transmtting out from or into the network. Users can establish a group of rules by certain messages in the packet, in which each rule to be applied on certain amount of matching messages: permit or deny. The users can apply these rules to the entrance or exit of specified switch, with which data stream in certain direction on certain port would have to follow the specified ACL rules in-and-out the switch. Please refer to chapter “ACL Configuration”.

3. Ip-prefix list

The ip-prefix list acts similarly to acl while more flexible and more understandable. The match object of ip-prefix is the destination address messages field of routing messages when applied in routing messages filtering.

An ip-prefix is identified by prefix list name. Each prefix list may contain multiple items, each of which specifies a matching range of a network prefix type and identifies with a sequence-number which specifies the matching check order of ip-prefix.

In the process of matching, the switch check each items identified by sequence-number in ascending order and the filter will be passed once certain items is matched(without checking rest items)

4. Autonomic system path information access-list as-path

The autonomic system path information access-list as-path is only used in BGP. In the BGP routing messages packet there is an autonomic system path field (in which autonomic system path the routing messages passes through is recorded). As-path is specially for specifying matching conditions for autonomic system path field.

As for relevant as-path configurations, please refer to the ip as-path command in BGP configuration.

5. community-list

Community-list is only for BGP. There is a community property field in the BGP routing messages packet for identifying a community. The community list is for specifying matching conditions for Community-list field.

As for relevant Community-list configuration, please refer to the ip as-path command in BGP configuration

4.1.2.2 IP Routing Policy Configuration Task List

1. Define route-map
2. Define the match clause in route-map
3. Define the set clause in route-map
4. Define address prefix list

1. Define route-map

Command	Explanation
Global mode	
route-map ***<map_name>*** **{deny | permit}** ***<sequence_num>*** **no route-map** ***<map_name>*** **[{deny | permit}** ***<sequence_num>*****]**	Configure route-map; the **no route-map** ***<map_name>*** **[{deny | permit}** ***<sequence_num>*****]** command deletes the route-map.

2. Define the match clause in route-map

Command	Explanation
Route-map configuration mode	

match as-path <*list-name*> **no match as-path [<*list-name*>]**	Match the autonomous system as path access-list the BGP route passes through; the **no match as-path [<*list-name*>]** command deletes match condition.
**match community <*community-list-name*	*community-list-num* > [exact-match]** **no match community [<*community-list-name*	*community-list-num* > [exact-match]]**	Match a community property access-list. The **no match community [<*community-list-name*	*community-list-num* > [exact-match]]** command deletes match condition.
match interface <*interface-name* > **no match interface [<*interface-name* >]**	Match by ports; The **no match interface [<*interface-name* >]** command deletes match condition.
**match ip <address	next-hop> <*ip-acl-name*	*ip-acl-num*	prefix-list *list-name*>** **no match ip <address	next-hop> [<*ip-acl-name*	*ip-acl-num*	prefix-list [*list-name*]>]**	Match the address or next-hop; The **no match ip <address	next-hop> [<*ip-acl-name*	*ip-acl-num*	prefix-list [*list-name*]>]** command deletes match condition.
match metric <*metric-val* > **no match metric [<*metric-val* >]**	Match the routing metric value; The **no match metric [<*metric-val* >]** command deletes match condition.
**match origin <egp	igp	incomplete >** **no match origin [<egp	igp	incomplete >]**	Match the route origin; The **no match origin [<egp	igp	incomplete >]** command deletes match condition.
**match route-type external <type-1	type-2 >** **no match route-type external [<type-1	type-2 >]**	Match the route type; The **no match route-type external [<type-1	type-2 >]** command deletes match condition.
match tag <*tag-val* > **no match tag [<*tag-val* >]**	Match the route tag; The **no match tag [<*tag-val* >]** command deletes match condition.

3. Define the set clause in route-map

Command	Explanation
Route-map configuration mode	
set aggregator as <*as-number*> <*ip_addr*> **no set aggregator as [<*as-number*> <*ip_addr*>]**	Distribute an AS No. for BGP aggregator; The no command deletes the configuration
set as-path prepend <*as-num*> **no set as-path prepend [<*as-num*>]**	Add a specified AS No. before the BGP routing messages as-path series; The no command deletes the configuration
set atomic-aggregate **no set atomic-aggregate**	Configure the BGP atomic aggregate property; The no command deletes the configuration
set comm-list <*community-list-name* | *community-list-num* > delete **no set comm-list <*community-list-name* | *community-list-num* > delete**	Delete BGP community list value; The no command deletes the configuration
set community [*AA:NN*] [internet] [local-AS] [no-advertise] [no-export] [none] [additive] **no set community [*AA:NN*] [internet] [local-AS] [no-advertise] [no-export] [none] [additive]**	Configure BGP community list value; The no command deletes the configuration
set extcommunity <rt | soo> <*AA:NN*> **no set extcommunity <rt | soo> [<*AA:NN*>]**	Configure BGP extended community list property; The no command deletes the configuration
set ip next-hop <*ip_addr*> **no set ip next-hop [<*ip_addr*>]**	Set next-hop IP address; The no command deletes the configuration
set local-preference <*pre_val*> **no set local-preference [<*pre_val*>]**	Set local preference; The no command deletes the configuration
set metric < *+/- metric_val | metric_val*> **no set metric [*+/- metric_val | metric_val*]**	Set routing metric value; The no command deletes the configuration
set metric-type <type-1 | type-2> **no set metric-type [<type-1 | type-2>]**	Set OSPF metric type; The no command deletes the configuration
set origin <egp | igp | incomplete > **no set origin [<egp | igp | incomplete >]**	Set BGP routing origin; The no command deletes the configuration

set originator-id <*ip_addr*> **no set originator-id [<*ip_addr*>]**	Set routing originator ID; The no command deletes the configuration
set tag <*tag_val*> **no set tag [<*tag_val*>]**	Set OSPF routing tag value; The no command deletes the configuration
set vpnv4 next-hop <*ip_addr*> **no set vpnv4 next-hop [<*ip_addr*>]**	Set BGP VPNv4 next-hop address; the no command deletes the configuration
set weight < *weight_val*> **no set weight [<*weight_val*>]**	Set BGP routing weight; The no command deletes the configuration

4. Define address prefix list

Command	Explanation
Global mode	
ip prefix-list <*list_name*> description <*description*> **no ip prefix-list <*list_name*> description**	Describe the prefix list; The **no ip prefix-list <*list_name*> description** command deletes the configuration.
ip prefix-list <*list_name*> [seq <*sequence_number*>] <deny | permit> < any / *ip_addr/mask_length* [ge *min_prefix_len*] [le *max_prefix_len*]> **no ip prefix-list <*list_name*> [seq <*sequence_number*>] [<deny | permit> < any / *ip_addr/mask_length* [ge *min_prefix_len*] [le *max_prefix_len*]>]**	Set the prefix list; The **no ip prefix-list <*list_name*> [seq <*sequence_number*>] [<deny | permit> < any / *ip_addr/mask_length* [ge *min_prefix_len*] [le *max_prefix_len*]>]** command deletes the configuration.

4.1.2.3 Configuration Examples

The figure below shows a network consisting of four Layer 3 switches. This example demonstrates how to set the BGP as-path properties through route-map. BGP protocol is applied among the Layer 3 switches. As for switchC, the network 192.68.11.0/24 can be reached through two paths in which one is AS-PATH 1 by IBGP (going through SwitchD), the other one is AS-PATH 2 by EBGP (going through SwitchB). BGP selects the shortest path, so AS-PATH 1 is the preferred path. If the path 2 is wished, which is through EBGP path, we can add two extra AS path numbers into the AS-PATH messages from SwitchA to SwitchD so as to change the determination SwitchC take to 192.68.11.0/24.

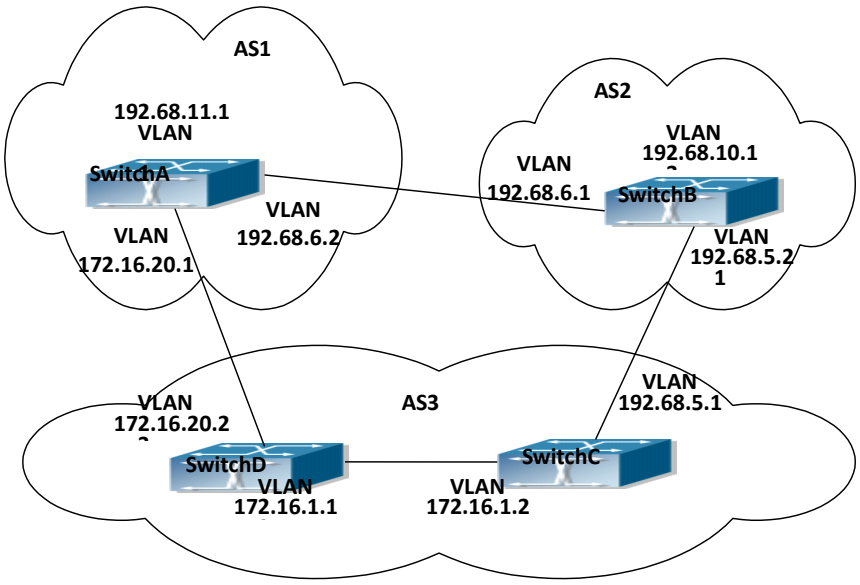


Figure 4-1 Policy routing Configuration

Configuration procedure: (only SwitchA is listed, configurations for other switches are omitted.)

The configuration of Layer 3 switchA:

SwitchA#config

SwitchA(config) #router bgp 1

SwitchA(config-router)#network 192.68.11.0 mask 255.255.255.0

SwitchA(config-router)#neighbor 172.16.20.2 remote-as 3

SwitchA(config-router)#neighbor 172.16.20.2 route-map AddAsNumbers out

SwitchA(config-router)#neighbor 192.68.6.1 remote-as 2

SwitchA(config-router)#exit

SwitchA(config)#route-map AddAsNumbers permit 10

SwitchA(config-route-map)#set as-path prepend 1 1

4.1.2.4 Troubleshooting

Faq: The routing protocol could not achieve the routing messages study under normal protocol running state

Troubleshooting: check following errors:

- ☞ Each node of route-map should at least has one node is permit match mode. When the route map is used in routing messages filtering, the routing messages will be considered not pass the routing messages filtering if certain routing messages does not pass the filtering of any nodes. When all nodes are set to deny mode, all routing messages will not pass the filtering in this route-map.
- ☞ Items in address prefix list should at least have one item set to permit mode. The deny mode items can be defined first to fast remove the unmatched routing messages, however if all the items are set to deny mode, any route will not be able to pass the filtering of this

address prefix list. We can define a permit 0.0.0.0/0 le 32 item after several deny mode items are defined so to permit all other routing messages pass through. Only default route will be matched in less-equal 32 is not specified.

4.2 Static Route

4.2.1 Introduction to Static Route

As mentioned earlier, the static route is the manually specified path to a network or a host. Static route is simple and consistent, and can prevent illegal route modification, and is convenient for load balance and route backup. However, it also has its own defects. Static route, as its name indicates, is static, it won't modify the route automatically on network failure, and manual configuration is required on such occasions, therefore it is not suitable for mid and large-scale networks.

Static route is mainly used in the following two conditions: 1) in stable networks to reduce load of route selection and routing data streams. For example, static route can be used in route to STUB network. 2) For route backup, configure static route in the backup line, with a lower priority than the main line.

Static route and dynamic route can coexist; layer3 switch will choose the route with the highest priority according to the priority of routing protocols. At the same time, static route can be introduced (redistribute) in dynamic route, and change the priority of the static route introduced as required.

4.2.2 Introduction to Default Route

Default route is a kind of static route, which is used only when no matching route is found. In the route table, default route in is indicated by a destination address of 0.0.0.0 and a network mask of 0.0.0.0, too. If the route table does not have the destination of a packet and has no default route configured, the packet will be discarded, and an ICMP packet will be sent to the source address indicate the destination address or network is unreachable.

4.2.3 Static Route Configuration Task List

1. Static route configuration
2. VRF configuration

1. Static route configuration

Command	Explanation
Global mode	

ip route {*<ip-prefix>* *<mask>* | *<ip-prefix>/<prefix-length>*} {*<gateway-address>* | *<gateway-interface>*} [*<distance>*] **no ip route {*<ip-prefix>* *<mask>* | *<ip-prefix>/<prefix-length>*} [*<gateway-address>* | *<gateway-interface>*] [*<distance>*]**	Set static routing; the **no ip route {*<ip-prefix>* *<mask>* | *<ip-prefix>/<prefix-length>*} [*<gateway-address>* | *<gateway-interface>*] [*<distance>*]** command deletes a static route entry

2. VRF configuration

Command	Explanation
Global mode	
ip route vrf <name> {<ip-prefix> <mask>|<ip-prefix/<prefix-length>} {<gateway-address>|<gateway-interface>} [<distance>] **no ip route vrf <name> {<ip-prefix> <mask>|<ip-prefix/<prefix-length>} [<gateway-address>|<gateway-interface>] [<distance>]**	Configure the static route, the no command will delete the static route.

4.2.4 Static Route Configuration Examples

The figure shown below is a simple network consisting of three layer3 switches, the network mask for all switches and PC is 255.255.255.0. PC-A and PC-C are connected via the static route set in SwtichA and SwitchC; PC3 and PC-B are connected via the static route set in SwtichC to SwitchB; PC-B and PC-C is connected via the default route set in SwitchB.

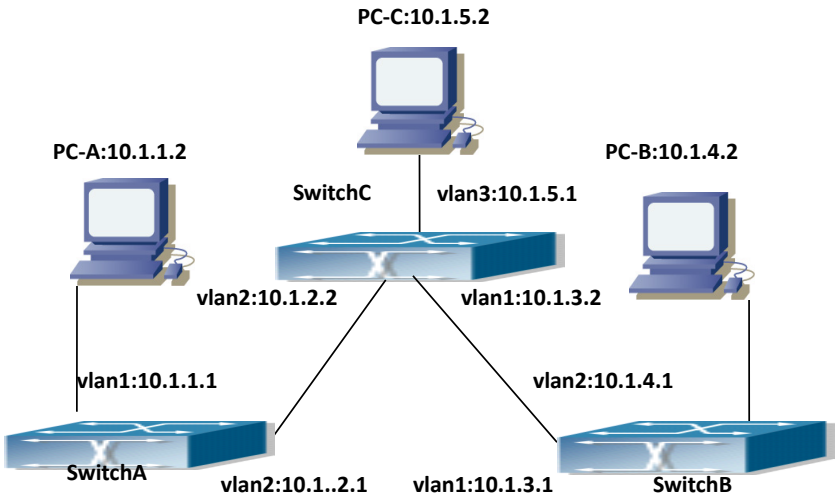


Figure 4-2 Static Route Configurations

Configuration steps:

Configuration of layer3 SwitchA

Switch#config

Switch (config) #ip route 10.1.5.0 255.255.255.0 10.1.2.2
Configuration of layer3 SwitchC
Switch#config
Next hop use the partner IP address
Switch(config)#ip route 10.1.1.0 255.255.255.0 10.1.2.1
Next hop use the partner IP address
Switch(config)#ip route 10.1.4.0 255.255.255.0 10.1.3.1
Configuration of layer3 SwitchB
Switch#config
Switch(config)#ip route 0.0.0.0 0.0.0.0 10.1.3.2
In this way, ping connectivity can be established between PC-A and PC-C, and PC-B and PC-C.

4.3 RIP

4.3.1 Introduction to RIP

RIP is first introduced in ARPANET, this is a protocol dedicated to small, simple networks. RIP is a distance vector routing protocol based on the Bellman-Ford algorithm. Network devices running vector routing protocol send two kind of information to the neighboring devices regularly:

• Number of hops to reach the destination network, or metrics to use or number of networks to pass.

• What is the next hop, or the director (vector) to use to reach the destination network.

The distance vector Layer 3 switch send all their route selecting tables to the neighbor layer3 switches at regular interval. A layer3 switch will build their own route selecting information table based on the information received from the neighbor layer3 switches. Then, it will send this information to its own neighbor layer3 switches. As a result, the route selection table is built on second hand information, route beyond 15 hops will be deemed as unreachable.

RIP protocol is an optional routing protocol based on UDP. Hosts using RIP send and receive packets on UDP port 520. All layer3 switches running RIP send their route table to all neighbor layer3 switches every 30 seconds for update. If no information from the partner is received in 180 seconds, then the device is deemed to have failed and the network connected to that device is considered to be unreachable. However, the route of that layer3 switch will be kept in the route table for another 120 seconds before deletion.

As layer3 switches running RIP built route table with second hand information, infinite count may occur. For a network running RIP routing protocol, when an RIP route becomes unreachable, the neighboring RIP layer3 switch will not send route update packets at once, instead, it waits until the update interval timeout (every 30 seconds) and sends the update packets containing that route. If before it receives the updated packet, its neighbors send packets containing the information about the failed neighbor, “infinite count” will be resulted. In other words, the route

of unreachable layer3 switch will be selected with the metrics increasing progressively. This greatly affects the route selection and route aggregation time.

To prevent “infinite count”, RIP provides mechanism such as “split horizon” and “triggered update” to solve route loop. “Split horizon” is done by avoiding sending to a gateway routes leaned from that gateway. There are two split horizon methods: “simple split horizon” and “poison reverse split horizon”. Simple split horizon deletes from the route to be sent to the neighbor gateways the routes learnt from the neighbor gateways; poison reverse split horizon not only deletes the abovementioned routes, but set the costs of those routes to infinite. “Triggering update” mechanism defines whenever route metric changed by the gateway, the gateway advertise the update packets immediately, regardless of the 30 second update timer status.

There two versions of RIP, version 1 and version 2. RFC1058 introduces RIP-I protocol, RFC2453 introduces RIP-II, which is compatible with RFC1723 and RFC1388. RIP-I updates packets by packets broadcast, subnet mask and authentication is not supported. Some fields in the RIP-I packets are not used and are required to be all 0’s; for this reason, such all 0's fields should be checked when using RIP-I, the RIP-I packets should be discarded if such fields are non-zero. RIP-II is a more improved version than RIP-I. RIP-II sends route update packets by multicast packets (multicast address is 224.0.0.9). Subnet mask field and RIP authentication filed (simple plaintext password and MD5 password authentication are supported), and support variable length subnet mask. RIP-II used some of the zero field of RIP-I and require no zero field verification. switch send RIP-II packets in multicast by default, both RIP-I and RIP-II packets will be accepted.

Each layer3 switch running RIP has a route database, which contains all route entries for reachable destination, and route table is built based on this database. When a RIP layer3 switch sent route update packets to its neighbor devices, the complete route table is included in the packets. Therefore, in a large network, routing data to be transferred and processed for each layer3 switch is quite large, causing degraded network performance.

Besides the above mentioned, RIP protocol allows route information discovered by the other routing protocols to be introduced to the route table. It can also be as the protocol exchanging route messages with CE on PE routers, and supports the VPN route/transmitting examples.

The operation of RIP protocol is shown below:

1. Enable RIP. The switch sends request packets to the neighbor layer3 switches by broadcasting; on receiving the request, the neighbor devices reply with the packets containing their local routing information.
2. The Layer3 switch modifies its local route table on receiving the reply packets and sends triggered update packets to the neighbor devices to advertise route update information. On receiving the triggered update packet, the neighbor lay3 switches send triggered update packets to their neighbor lay3 switches. After a sequence of triggered update packet broadcast, all layer3 switches get and maintain the latest route information.

In addition, RIP layer3 switches will advertise its local route table to their neighbor devices every 30 seconds. On receiving the packets, neighbor devices maintain their local route table, select the best route and advertise the updated information to their own neighbor devices, so that the updated routes are globally valid. Moreover, RIP uses a timeout mechanism for outdated route, that is, if a switch does not receive regular update packets from a neighbor within a certain

interval (invalid timer interval), it considers the route from that neighbor invalid, after holding the route fro a certain interval (holddown timer interval), it will delete that route.

4.3.2 RIP Configuration Task List

1. Enable RIP (required)
   (1) Enable/disable RIP module.
   (2) Enable interface to send/receive RIP packets
2. Configure RIP protocol parameters (optional)
   (1) Configure RIP sending mechanism
      1) Configure specified RIP packets transmission address
      2) Configure RIP interface broadcast
   (2) Configure the RIP routing parameters
      1) Configure route introduction (default route metric, configure routes of the other protocols to be introduced in RIP)
      2) Configure interface authentication mode and password
      3) Configure the route deviation
      4) Configure and apply route filter
      5) Configure Split Horizon
   (3) Configure other RIP protocol parameters
      1) Configure the managing distance of RIP route
      2) Configure the RIP route capacity limit in route table
      3) Configure the RIP update, timeout, holddown and other timer.
      4) Configure the receiving buffer size of RIP UDP
3. Configure RIP-I/RIP-II switch
   (1) Configure the RIP version to be used in all interfaces
   (2) Configure the RIP version to send/receive in all interfaces
   (3) Configure whether to enable RIP packets sending/receiving for interfaces
4. Delete the specified route in RIP route table
5. Configure the RIP routing aggregation
   (1) Configure aggregation route of IPv4 route mode
   (2) Configure aggregation route of IPv4 interface configuration mode
   (3) Display IPv4 aggregation route information
6. Configure redistribution of OSPF routing to RIP
   (1) Enable Redistribution of OSPF routing to RIP
   (2) Display and debug the information about configuration of redistribution of OSPF routing to RIP

1. Enable RIP protocol

Applying RIP route protocol with basic configuration in switch is simple. Normally you only have to open the RIP switch and configure the segments running RIP, namely send and receive the RIP data packet by default RIP configuration. The version of data packet sending and receiving

is variable when needed, allow/deny sending, receiving RIP data packet. Refer to 3.

Command	Explanation
Global Mode	
router rip **no router rip**	Enables RIP; the **no router rip** command disables RIP.
Router and address family configuration mode	
network *<A.B.C.D/M | ifname|vlan>* **no network *<A.B.C.D/M | ifname|vlan>***	Enables the segment running RIP protocol; the **no network *<A.B.C.D/M | ifname|vlan>*** command deletes the segment.

2. Configure RIP protocol parameters

（1）Configure RIP packet transmitting mechanism

1）Configure the RIP data packet point-transmitting

2）Configure the Rip broadcast

Command	Explanation
Router Configuration Mode	
neighbor *<A.B.C.D>* **no neighbor *<A.B.C.D>***	Specify the IP address of the neighbor router needs point-transmitting; the **no neighbor *<A.B.C.D>*** command cancels the appointed router.
passive-interface*<ifname|vlan>* **no passive-interface*<ifname|vlan >***	Block the RIP broadcast on specified pot and the RIP data packet is only transmittable among Layer 3 switch configured with neighbor. The **no passive-interface*<ifname|vlan >*** command cancels the function.

（2）Configure RIP route parameters

1）Configure route introduction (default route metric, configure routes of the other protocols to be introduced in RIP)

Command	Explanation
Router Configuration Mode	
default-metric <*value*> **no default-metric**	Sets the default route metric for route to be introduced; the **no default-metric** command restores the default setting.
redistribute {kernel |connected| static| ospf | isis| bgp} [metric<*value*>] [route-map<*word*>] **no redistribute {kernel |connected| static| ospf | isis| bgp} [metric<*value*>] [route-map<*word*>]**	Redistribute the routes distributed in other routing protocols into the RIP data packet; the **no redistribute {kernel |connected| static| ospf | isis| bgp} [metric<value>] [route-map<word>]** command cancels the distributed route of corresponding protocols.
default-information originate **no default-information originate**	Generate a default route to the RIP protocol; the **no default-information originate** command cancels the feature.

2）Configure interface authentication mode and password

Command	Explanation
Interface configuration mode	
ip rip authentication mode { text| md5} **no ip rip authentication mode [text| md5]**	Sets the authentication method; the **no ip rip authentication mode [text| md5]** command cancels the authentication action.
ip rip authentication string *<text>* **no ip rip authentication string**	Sets the authentication key; the **no ip rip authentication string** command means no key is needed.
ip rip authentication key-chain *<name-of-chain>* **no ip rip authentication key-chain [*<name-of-chain>*]**	Sets the key chain used in authentication, the **no ip rip authentication key-chain [*<name-of-chain>*]** command means the key chain is not used.
ip rip authentication cisco-compatible **no ip rip authentication cisco-compatible**	After configure this command, configure MD5 authentication, then can receive RIP packet of Cisco, the no command restores the default configuration.
Global mode	
key chain *<name-of-chain>* **no key chain *< name-of-chain >***	Enter keychain mode, and configure a key chain, the **no key chain *< name-of-chain >*** command deletes the key chain.
Keychain mode	
key *<keyid>* **no key *<keyid>***	Enter the keychain-key mode and configure a key of the keychain; the **no key *<keyid>*** command deletes one key.
Keychain-key mode	
key-string *<text>* **no key-string *<text>***	Configure the password used by the key, the **no key-string *<text>*** command deletes the password.
accept-lifetime *<start-time>* {*<end-time>*| duration*<seconds>*| infinite} **no accept-lifetime**	Configure a key on the key chain and accept it as an authorized time; the **no accept-lifetime** command deletes it.
send-lifetime *<start-time>* {*<end-time>*| duration*<seconds>*| infinite} **no send-lifetime**	Configure the transmitting period of a key on the key chain; the **no send-lifetime** command deletes the send-lifetime.

3）Configure the route deviation

Command	Explanation
Router configuration mode	

offset-list <access-list-number | access-list-name> {in | out } <number> [<ifname>] **no offset-list <access-list-number |access-list-name> {in|out }<number >[<ifname>]**	Configure that provide a deviation value to the route metric value when the port sends or receives RIP data packet; the **no offset-list <access-list-number |access-list-name> {in|out } <number >[<ifname>]** command removes the deviation table.

4）Configure and apply the route filtering

Command	Explanation
Router configuration mode	
distribute-list {< ***access-list-number /access-list-name >*****|prefix<*** ***prefix-list-name*** **>}{in|out} [<*** ***ifname*** **>]** **no distribute-list {<** ***access-list-number /access-list-name >*****|prefix<*** ***prefix-list-name*** **>}{in|out} [<*** ***ifname*** **>]**	Configure and apply the access table and prefix table to filter the routes. The **no distribute-list {<** ***access-list-number /access-list-name>*****|prefix<*** ***prefix-list-name*** **>}{in|out} [<*** ***ifname*** **>]** command means do not use the access table and prefix table.

5）Configure the split horizon

Command	Explanation
Interface configuration mode	
ip rip split-horizon [poisoned] **no ip rip split-horizon**	Configure that take the split horizon when the port sends data packets; poisoned for poison reverse the **no ip rip split-horizon** command cancels the split horizon.

（3）Configure other RIP protocol parameters

1）Configure RIP routing priority

2）Configure the RIP route capacity limit in route table

3）Configure timer for RIP update, timeout and hold-down

4）Configure RIP UDP receiving buffer size

Command	Explanation
Router configuration mode	
distance <* ***number*** **> [<*** ***A.B.C.D/M*** **>] [<*** ***access-list-name|access-list-number*** **>]** **no distance [<*** ***A.B.C.D/M*** **>]**	Specify the route administratively distance of RIP protocol; the **no distance [<*** ***A.B.C.D/M*** **>]** command restores the default value 120.
maximum-prefix <* ***maximum-prefix*** **>[<*** ***threshold*** **>]** **no maximum-prefix <*** ***maximum-prefix*** **>** **no maximum-prefix**	Configure the maximum of RIP route; the **no maximum-prefix <*** ***maximum-prefix*** **>** **no maximum-prefix** command cancels the limit.
timers basic <* ***update*** **> <*** ***invalid*** **> <*** ***garbage*** **>** **no timers basic**	Adjust the update, timeout and garbage collection time, the **no timers basic** command restores the default configuration.

recv-buffer-size <*size*> **no recv-buffer-size**	The command configures the UDP receiving buffer size of the RIP; the **no recv-buffer-size** command restores the system default values.

3. Configure RIP-I/RIP-II toggling

（1）Configure the RIP version to be used in all ports

Command	Explanation
RIP configuration mode	
version { 1 | 2 } **no version**	Configure the versions of all the RIP data packets transmitted/received by the Layer 3 switch port sending/receiving the **no version** command restores the default configuration, version 2.

（2）Configure the RIP version to send/receive in all ports.

（3）Configure whether to enable RIP packets sending/receiving for ports

Command	Explanation
Interface configuration mode	
ip rip send version { 1 | 1-compatible | 2 } **no ip rip send version**	Sets the version of RIP packets to send on all ports; the **no ip rip send version** command set the version to the one configured by the version command.
ip rip receive version {1 | 2 | } **no ip rip receive version**	Sets the version of RIP packets to receive on all ports; the no action of this command set the version to the one configured by the version command.
ip rip receive-packet **no ip rip receive-packet**	Enables receiving RIP packets on the interface; the **no ip rip receive-packet** command close data receiving on this port.
ip rip send-packet **no ip rip send-packet**	Enables sending RIP packets on the interface; the **no ip rip send-packet** command disables sending RIP packets on the interface.

4. Delete the specified route in RIP route table

Command	Explanation
Admin Mode	
clear ip rip route {<*A.B.C.D/M*>|kernel|static|connected|rip|ospf|isis|bgp|all}	The command deletes a specified route from the RIP route table.

5. Configure the RIP routing aggregation

（1） Configure IPv4 aggregation route globally

Command	Explanation
Router Configuration Mode	
ip rip aggregate-address A.B.C.D/M **no ip rip aggregate-address A.B.C.D/M**	To configure or delete IPv4 aggregation route globally.

（2） Configure IPv4 aggregation route on interface

Command	Explanation
Interface Configuration Mode	
ip rip aggregate-address A.B.C.D/M **no ip rip aggregate-address A.B.C.D/M**	To configure or delete IPv4 aggregation route on interface.

（3） Display IPv4 aggregation route information

Command	Explanation
Admin Mode and Configuration Mode	
show ip rip aggregate	To display aggregation route information.

6. Configure redistribution of OSPF routing to RIP

(1) Enable Redistribution of OSPF routing to RIP

Command	Explanation
Router RIP Configuration Mode	
redistribute ospf [<*process-id*>] [metric <*value*>] [route-map <*word*>] **no redistribute ospf [<process-id>]**	To enable or disable the redistribution of OSPF routing to RIP.

(2) Display and debug the information about configuration of redistribution of OSPF routing to RIP

Command	Explanation
Admin Mode and Configuration Mode	
show ip rip redistribute	To display the information about configuration of redistribute from other routing.
Admin Mode	
debug rip redistribute message send **no debug rip redistribute message send** **debug rip redistribute route receive** **no debug rip redistribute route receive**	To enable or disable debugging messages sent by RIP for redistribution of OSPF routing. To enable or disable debugging messages received from NSM.

4.3.3 RIP Examples

4.3.3.1 Typical RIP Examples

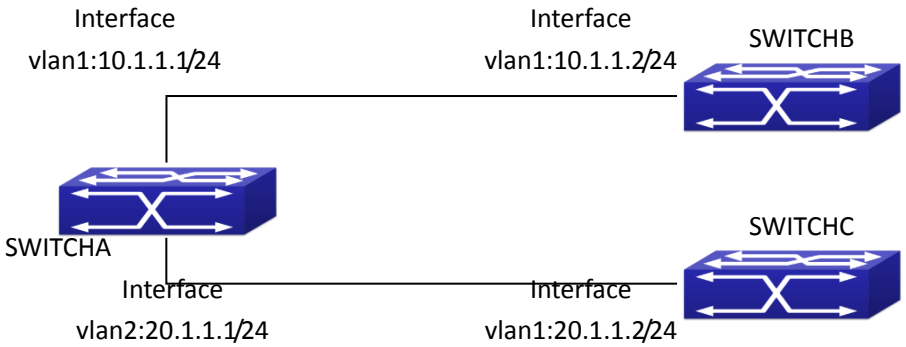


Figure 4-3 RIP example

In the figure shown above, a network consists of three Layer 3 switches, in which SwitchA connected with SwitchB and SwitchC, and RIP routing protocol is running in all of the three switches. SwitchA (interface vlan1：10.1.1.1,interface vlan2：20.1.1.1) exchanges Layer 3 switch update messages only with SwitchB (interface vlan1：10.1.1.2), but not with SwitchC (interface vlan 2: 20.1.1.2).

SwitchA, SwitchB, SwitchC configurations are as follows:

a) Layer 3 SwitchA：

Configure the IP address of interface vlan 1
SwitchA#config
SwitchA(config)# interface vlan 1
SwitchA(Config-if-Vlan1)# ip address 10.1.1.1 255.255.255.0
SwitchA(config-if-Vlan1)#
Configure the IP address of interface vlan 2
SwitchA(config)# vlan 2
SwitchA(Config-Vlan2)# switchport interface ethernet 1/0/2
Set the port Ethernet1/0/2 access vlan 2 successfully
SwitchA(Config-Vlan2)# exit
SwitchA(config)# interface vlan 2
SwitchA(Config-if-Vlan2)# ip address 20.1.1.1 255.255.255.0
Initiate RIP protocol and configure the RIP segments
SwitchA(config)#router rip
SwitchA(config-router)#network vlan 1
SwitchA(config-router)#network vlan 2
SwitchA(config-router)#exit
Configure that the interface vlan 2 do not transmit RIP messages to SwitchC
SwitchA(config)#router rip
SwitchA(config-router)#passive-interface vlan 2
SwitchA(config-router)#exit
SwitchA(config) #

b) Layer 3 SwitchB

Configure the IP address of interface vlan 1
SwitchB#config
SwitchB(config)# interface vlan 1
SwitchB(Config-if-Vlan1)# ip address 10.1.1.2 255.255.255.0
SwitchB(Config-if-Vlan1)exit
Initiate RIP protocol and configure the RIP segments
SwitchB(config)#router rip
SwitchB(config-router)#network vlan 1
SwitchB(config-router)#exit

c) Layer 3 SwitchC

SwitchC#config
SwitchC(config)# interface vlan 1
Configure the IP address of interface vlan 1

SwitchC(Config-if-Vlan1)# ip address 20.1.1.2 255.255.255.0
SwitchC(Config-if-Vlan1)#exit
Initiate RIP protocol and configure the RIP segments
SwitchC(config)#router rip
SwitchC(config-router)#network vlan 1
SwitchC(config-router)#exit

4.3.3.2 Typical Examples of RIP aggregation function

The application topology as follows：

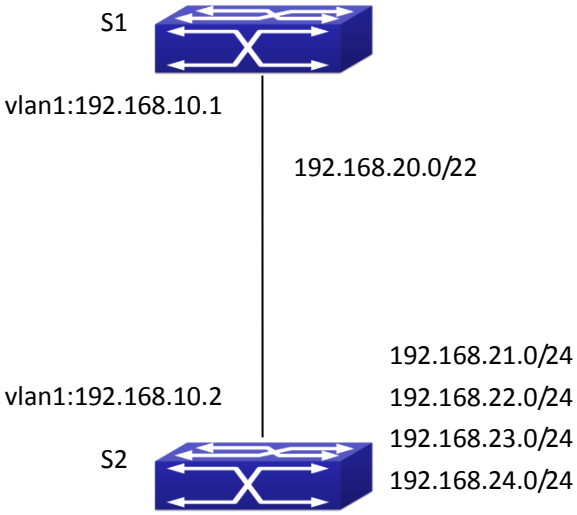


Figure 4-4 Typical application of RIP aggregation

As the above network topology, S2 is connected to S1 through interface vlan1, there are other 4 subnet routers of S2, which are 192.168.21.0/24, 192.168.22.0/24, 192.168.23.0/24, 192.168.24.0/24. S2 supports route aggregation, and to configure aggregation route 192.168.20.0/22 in interface vlan1 of S2, after that, sending router messages to S1 through vlan1, and put the four subnet routers aggregated to one router as 192.168.20.0/22, and send to S1, and not send subnet to neighbor. It can reduce the router table of S1, save the memory.

S1 configuration list:
S1(config)#router rip
S1(config-router) #network vlan 1
S2 configuration list:
S2(config)#router rip
S2(config-router) #network vlan 1
S2(config-router) #exit
S2(config)#in vlan 1
S2(Config-if-Vlan1)# ip rip agg 192.168.20.0/22

4.3.4 RIP Troubleshooting

The RIP protocol may not be working properly due to errors such as physical connection,

configuration error when configuring and using the RIP protocol. So users should pay attention to following:

- ☞ First ensure the physic connection is correct
- ☞ Second, ensure the interface and chain protocol are UP (use **show interface** command)
- ☞ Then initiate the RIP protocol (use **router rip** command) and configure the segment (use **network** command) and set RIP protocol parameter on corresponding interfaces, such as the option between RIP-I and RIP-II
- ☞ After that, one feature of RIP protocol should be noticed ---the Layer 3 switch running RIP protocol sending route updating messages to all neighboring Layer 3 switches every 30 seconds. A Layer 3 switch is considered inaccessible if no route updating messages from the switch is received within 180 seconds, then the route to the switch will remains in the route table for 120 seconds before it is deleted. Therefore, if to delete a RIP route, this route item is assured to be deleted from route table after 300 seconds.
- ☞ When exchanging routing messages with CE using RIP protocol on the PE router, we should first create corresponding VPN routing/transmitting examples to associate with corresponding interfaces. Then enter the RIP address family mode configuring corresponding parameters. If the RIP routing problem remains unresolved, please use debug rip command to record the debug message in three minutes, and send them to our technical service center.

4.4 OSPF

4.4.1 Introduction to OSPF

OSPF is abbreviation for Open Shortest Path First. It is an interior dynamic routing protocol for autonomous system based on link-state. The protocol creates a link-state database by exchanging link-states among layer3 switches, and then uses the Shortest Path First algorithm to generate a route table basing on that database.

Autonomous system (AS) is a self-managed interconnected network. In large networks, such as the Internet, a giant interconnected network is broken down to autonomous systems. Big enterprise networks connecting to the Internet are independent AS, since the other hosts on the Internet are not managed by those AS and they don’t share interior routing information with the layer3 switches on the Internet.

Each link-state Layer3 switch can provide information about the topology with its neighboring Layer3 switches.

• The network segment (link) connecting to the layer3 switch

• State of the connecting link

Link-state information is flooded throughout the network so that all Layer3 switches can get firsthand information. Link-state Layer3 switches will not broadcast all information contained in their route tables; instead, they only send changed link-state information. Link-state Layer3 switches establish neighborhood by sending “HELLO” to their neighbors, then link-state advertisements (LSA) will be sent among neighboring Layer3 switches. Neighboring Layer3 switch copy the LSA to their routing table and transfer the information to the rest part of the network. This process is referred to as “flooding”. In this way, firsthand information is sent throughout the network to provide accurate map for creating and updating routes in the network. Link-state routing protocols use cost instead of hops to decide the route. Cost is assigned automatically or manually. According to the algorithm in link-state protocol, cost can be used to calculate the hop number for packets to pass, link bandwidth, and current load of the link. The administrator can even add weight for better assessment of the link-state.

1) When a link-state layer3 switch enters a link-state interconnected network, it sends a HELLO packet to get to know its neighbors and establish neighborhood.

2) The neighbors respond with information about the links they are connecting and the related costs.

3) The originate layer3 switch uses this information to build its own routing table

4) Then, as part of the regular update, layer3 switch send link-state advertisement (LSA) packets to its neighboring layer3 switches. The LSA include links and related costs of that layer3 switch.

5) Each neighboring layer3 switch copies the LSA packet and passes it to the next neighbor (i.e. flooding).

6) Since routing database is not recalculated before layer3 switch forwards LSA flooding, the converging time is greatly reduced.

One major advantage of link-state routing protocols is the fact that infinite counting is impossible, this is because of the way link-state routing protocols build up their routing table. The second advantage is that converging in a link-state interconnected network is very fast, once the routing topology changes, updates will be flooded throughout the network very soon. Those advantages release some layer3 switch resources, as the process ability and bandwidth used by bad route information are minor.

The features of OSPF protocol include the following: OSPF supports networks of various scales, several hundreds of layer3 switches can be supported in an OSPF network. Routing topology changes can be quickly found and updating LSAs can be sent immediately, so that routes converge quickly. Link-state information is used in shortest path algorithm for route calculation, eliminating loop route. OSPF divides the autonomous system into areas, reducing database size, bandwidth occupation and calculation load. (According to the position of layer3 switches in the autonomous system, they can be grouped as internal area switches, area border switches, AS border switches and backbone switches). OSPF supports load balance and multiple routes to the same destination of equal costs. OSPF supports 4 level routing mechanisms (process routing according to the order of intra-area path, inter-area path, type 1 external path and type 2 external path). OSPF supports IP subnet and redistribution of routes from the other routing protocols, and interface-based packet verification. OSPF supports sending packets in multicast.

Each OSPF layer3 switch maintains a database describing the topology of the whole

autonomous system. Each layer3 switch gathers the local status information, such as available interface, reachable neighbors, and sends link-state advertisement (sending out link-state information) to exchange link-state information with other OSPF layer3 switches to form a link-state database describing the whole autonomous system. Each layer3 switch builds a shortest path tree rooted by itself according to the link-state database, this tree provides the routes to all nodes in an autonomous system. If two or more layer3 switches exist (i.e. multi-access network), "designated layer3 switch" and "backup designated layer3 switch" will be selected. Designated layer3 switch is responsible for spreading link-state of the network. This concept helps reducing the traffic among the Layer3 switches in multi-access network.

OSPF protocol requires the autonomous system to be divided into areas. That is to divide the autonomous system into 0 area (backbone area) and non-0 areas. Routing information between areas are further abstracted and summarized to reduce the bandwidth required in the network. OSPF uses four different kinds of routes; they are intra-area route, inter-area route, type 1 external route and type 2 external route, in the order of highest priority to lowest. The route inside an area and between areas describes the internal network structure of an autonomous system, while external routes describe how to select the routing information to destination outside the autonomous system. The first type of exterior route corresponds to the information introduced by OSPF from the other interior routing protocols, the costs of those routes are comparable with the costs of OSPF routes; the second type of exterior route corresponds to the information introduced by OSPF from the other exterior routing protocols, but the costs of those routes are far greater than that of OSPF routes, so OSPF route cost is ignored when calculating route costs.

OSPF areas are centered with the Backbone area, identified as Area 0, all the other areas must be connected to Area 0 logically, and Area 0 must be continuous. For this reason, the concept of virtual link is introduced to the backbone area, so that physically separated areas still have logical connectivity to the backbone area. The configurations of all the layer3 switches in the same area must be the same.

In conclusion, LSA can only be transferred between neighboring Layer3 switches, OSPF protocol includes 5 types of LSA: router LSA, network LSA, network summary LSA to the other areas, ASBR summary LSA and AS external LSA. They can also be called type1 LSA, type2 LSA, type3 LSA, type4 LSA, and type5 LSA. Router LSA is generated by each layer3 switch inside an OSPF area, and is sent to all the other neighboring layer3 switches in the same area; network LSA is generated by the designated layer3 switch in the OSPF area of multi-access network, and is sent to all other neighboring layer3 switches in this area. (In order to reduce traffic on layer3 switches in the multi-access network, "designated layer3 switch" and "backup designated layer3 switch" should be selected in the multi-access network, and the network link-state is broadcasted by the designated layer3 switch); network summary LSA is generated by border switches in an OSPF area , and is transferred among area border layer3 switches; AS external LSA is generated by layer3 switches on external border of AS, and is transferred throughout the AS.

As to autonomous systems mainly advertises exterior link-state, OSPF allow some areas to be configured as STUB areas to reduce the size of the topology database. Type4 LSA (ASBR summary LSA) and type5 LSA (AS external LSA) are not allowed to flood into/through STUB areas. STUB areas must use the default routes, the layer3 switches on STUB area edge advertise the default routes to STUB areas by type 3 summary LSA, those default routes only floods inside STUB

area and will not get out of STUB area. Each STUB area has a corresponding default route, the route from a STUB area to AS exterior destination must rely on the default route of that area.

The following simply outlines the route calculation process of OSPF protocol:

1） Each OSPF-enabled layer3 switch maintains a database (LS database) describing the link-state of the topology structure of the whole autonomous system. Each layer3 switch generates a link-state advertisement according to its surrounding network topology structure (router LSA), and sends the LSA to other layer3 switches through link-state update (LSU) packets. Thus each layer3 switches receives LSAs from other layer3 switches, and all LSAs are combined to the link-state database.

2） Since a LSA is the description of the network topology structure around a layer3 switch, the LS database is the description of the network topology structure of the whole network. The layer3 switches can easily create a weighted vector map according to the LS database. Obviously, all layer3 switches in the same autonomous system will have the same network topology map.

3） Each layer3 switch uses the shortest path first (SPF) algorithm to calculate a tree of shortest path rooted by itself. The tree provides the route to all the nodes in the autonomous system, leaf nodes consist of the exterior route information. The exterior route can be marked by the layer3 switch broadcast it, so that additional information about the autonomous system can be recorded. As a result, the route table of each layer3 switch is different.

OSPF protocol is developed by the IETF; the OSPF v2 widely used now is fulfilled according to the content described in RFC2328.

4.4.2 OSPF Configuration Task List

The OSPF configuration may be different from the configuration procedure to switches of the other manufacturers. It is a two-step process:

1、 Enable OSPF in the Global Mode; 2、Configure OSPF area for the interfaces. The configuration task list is as follows:

1. Enable OSPF protocol (required)
   (1) Enable/disable OSPF protocol (required)
   (2) Configure the ID number of the layer3 switch running OSPF (optional)
   (3) Configure the network scope for running OSPF (optional)
   (4) Configure the area for the interface (required)
2. Configure OSPF protocol parameters (optional)
   (1) Configure OSPF packet sending mechanism parameters
      1) Configure OSPF packet verification
      2) Set the OSPF interface to receive only
      3) Configure the cost for sending packets from the interface

4) Configure OSPF packet sending timer parameter (timer of broadcast interface sending HELLO packet to poll, timer of neighboring layer3 switch invalid timeout, timer of LSA transmission delay and timer of LSA retransmission.

(2) Configure OSPF route introduction parameters

1) Configure default parameters (default type, default tag value, default cost)

2) Configure the routes of the other protocols to introduce to OSPF.

(3) Configure OSPF importing the routes of other OSPF processes

1) Enable the function of OSPF importing the routes of other OSPF processes

2) Display relative information

3) Debug

(4) Configure other OSPF protocol parameters

1) Configure OSPF routing protocol priority

2) Configure cost for OSPF STUB area and default route

3) Configure OSPF virtual link

4) Configure the priority of the interface when electing designated layer3 switch (DR).

5) Configure to keep a log for OSPF adjacency changes or not

6) Filter the route obtained by OSPF

3. Disable OSPF protocol

1. Enable OSPF protocol

Basic configuration of OSPF routing protocol on switch is quite simple, usually only enabling OSPF and configuration of the OSPF area for the interface are required. The OSPF protocol parameters can use the default settings. If OSPF protocol parameters need to be modified, please refer to “2. Configure OSPF protocol parameters”.

Command	Explanation
Global Mode	
[no] router ospf [process <*id*>]	Enables OSPF protocol; the **no** router ospf command disables OSPF protocol.（required)
OSPF Protocol Configuration Mode	
router-id <*router_id*> no **router-id**	Configures the ID number for the layer3 switch running OSPF; the **no router id** command cancels the ID number. The IP address of an interface is selected to be the layer3 switch ID. (optional)
[no] network {***<network> <mask> / <network>/<prefix>***} **area <*area_id*>**	Configure certain segment to certain area, the **no network** {<network> ***<mask> / <network>/<prefix>}*** **area <area_*id*> *command*** cancels this configuration. (required)

2. Configure OSPF protocol parameters

（1）Configure OSPF packet sending mechanism parameters

1）Configure OSPF packet verification

2）Set the OSPF interface to receive only

3）Configure the cost for sending packets from the interface

Command	Explanation
Interface Configuration Mode	
ip ospf authentication { message-digest | null} **no ip ospf authentication**	Configures the authentication method by the interface to accept OSPF packets; the **no ip ospf authentication** command restores the default settings.
ip ospf [<*ip-address*>] authentication-key <*0 LINE | 7 WORD | LINE*> **no ip ospf [<*ip-address*>] authentication**	Specify the authentication key required in sending and receiving OSPF packet on the interface; the no command cancels the authentication key.
[no] passive-interface <*ifname*> [<*ip-address*>]	Sets an interface to receive only, the **no passive-interface <*ifname*>[<*ip-address*>]** command cancels this configuration.
ip ospf cost <*cost* > **no ip ospf cost**	Sets the cost for running OSPF on the interface; the **no ip ospf cost** command restores the default setting.

4) Configure OSPF packet sending timer parameter (timer of broadcast interface sending HELLO packet to poll, timer of neighboring layer3 switch invalid timeout, timer of LSA transmission delay and timer of LSA retransmission).

Command	Explanation
Interface Configuration Mode	
ip ospf hello-interval <*time*> **no ip ospf hello-interval**	Sets interval for sending HELLO packets; the **no ip ospf hello-interval** command restores the default setting.
ip ospf dead-interval <*time* > **no ip ospf dead-interval**	Sets the interval before regarding a neighbor layer3 switch invalid; the **no ip ospf dead-interval** command restores the default setting.
ip ospf transit-delay <*time*> **no ip ospf transit-delay**	Sets the delay time before sending link-state broadcast; the **no ip ospf transmit-delay** command restores the default setting.
ip ospf retransmit <*time*> **no ip ospf retransmit**	Sets the interval for retransmission of link-state advertisement among neighbor layer3 switches; the **no ip ospf retransmit** command restores the default setting.

（2）Configure OSPF route introduction parameters

Configure the routes of the other protocols to introduce to OSPF.

Command	Explanation
OSPF Protocol Configuration Mode	
redistribute { bgp | connected | static | rip | kernel} [metric-type { 1 | 2 }] [tag <*tag*>] [metric <*cost_value*>] [router-map <*WORD*>] **no redistribute { bgp | connected | static | rip | kernel }**	Distribute other protocols to find routing and static routings as external routing messages the **no redistribute {bgp | connected | static | rip | kernel}** command cancels the distributed external messages.

（3）Configure OSPF importing the routes of other OSPF processes

1）Enable the function of OSPF importing the routes of other OSPF processes

Command	Explanation
Router OSPF Mode	
redistribute ospf [<*process-id*>] [metric<*value*>] [metric-type {1|2}][route-map<*word*>] **no redistribute ospf [<*process-id*>] [metric<*value*>] [metric-type {1|2}][route-map<*word*>]**	Enable or disable the function of OSPF importing the routes of other OSPF processes.

2）Display relative information

Command	Explanation
Admin Mode or Configure Mode	
show ip ospf [<*process-id*>] redistribute	Display the configuration information of the OSPF process importing other outside routes.

3）Debug

Command	Explanation
Admin Mode	
debug ospf redistribute message send **no debug ospf redistribute message send** **debug ospf redistribute route receive** **no debug ospf redistribute route receive**	Enable or disable debugging of sending command from OSPF process redistributed to other OSPF process routing. Enable or disable debugging of received routing message from NSM for OSPF process.

（4）Configure other OSPF protocol parameters

1）Configure how to calculate OSPF SPF algorithm time

2）Configure the LSA limit in the OSPF link state database

3）Configure various OSPF parameters

Command	Explanation
OSPF Protocol Configuration Mode	

timers spf <*interval*> **no timers spf**	Configure the SPF timer of OSPF; the **no timers spf** command restores the default settings.
overflow database {<*max-LSA*> [hard | soft] | external <*max-LSA*> <*recover time*>} **no overflow database [external <*max-LSA* > < *recover time* >]**	Configure the LSA limit in current OSPF process database; the **no overflow database [external < *max-LSA* > < *recover time* >]** command restores the default settings.
area <id> {authentication [message-digest] | default-cost <*cost*> | filter-list {access | prefix} <*WORD*> {in | out} | nssa [default-information-originate | no-redistribution | no-summary | translator-role] | range <*range*> | stub [no-summary] | virtual-link <*neighbor*>} **no area <id> {authentication | default-cost | filter-list {access | prefix} <*WORD*> {in | out} | nssa [default-information-originate | no-redistribution | no-summary | translator-role] | range <*range*> | stub [no-summary] | virtual-link <*neighbor*>}**	Configure the parameters in OSPF area (STUB area, NSSA area and virtual links); the **no area <id> {authentication | default-cost | filter-list {access | prefix} <*WORD*> {in | out} | nssa [default-information-originate | no-redistribution | no-summary | translator-role] | range <*range*> | stub [no-summary] | virtual-link <*neighbor*>}** command restores the default settings.

4）Configure the priority of the interface when electing designated layer3 switch (DR).

Command	Explanation
Interface Configuration Mode	
ip ospf priority <*priority*> **no ip ospf priority**	Sets the priority of the interface in “designated layer3 switch” election; the **no ip ospf priority** command restores the default setting.

5）Configure to keep a log for OSPF adjacency changes or not

Command	Explanation
OSPF Protocol Configuration Mode	
log-adjacency-changes detail **no log-adjacency-changes detail**	Configure to keep a log for OSPF adjacency changes or not.

6）Filter the route obtained by OSPF

Command	Explanation
OSPF Protocol Configuration Mode	
filter-policy <access-list-name> **no filter-policy**	Use access list to filter the route obtained by OSPF, the no command cancels the route filtering.

3. Disable OSPF protocol

Command	Explanation
Global Mode	

no router ospf [process *<id>*]	Disables OSPF routing protocol.

4.4.3 OSPF Examples

4.4.3.1 Configuration Example of OSPF

Scenario 1: OSPF autonomous system.

This scenario takes an OSPF autonomous system consists of five switch for example.

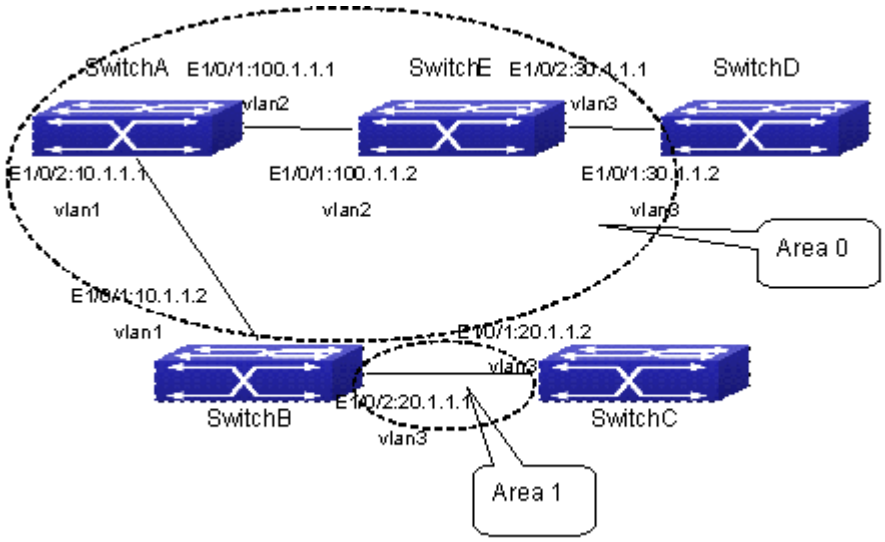


Figure 4-5 Network topology of OSPF autonomous system

The configuration for layer3 Switch1 and Switch5 is shown below:

Layer 3 Switch1

Configuration of the IP address for interface vlan1

Switch1#config

Switch1(config)# interface vlan 1

Switch1(config-if-vlan1)# ip address 10.1.1.1 255.255.255.0

Switch1(config-if-vlan1)#exit

Configuration of the IP address for interface vlan2

Configure the IP address of interface vlan2

Switch1(config)# interface vlan 2

Switch1(config-if-vlan2)# ip address 100.1.1.1 255.255.255.0

Switch1 (config-if-vlan2)#exit

Enable OSPF protocol, configure the area number for interface vlan1 and vlan2.

Switch1(config)#router ospf

Switch1(config-router)#network 10.1.1.0/24 area 0

Switch1(config-router)#network 100.1.1.0/24 area 0

Switch1(config-router)#exit

Switch1(config)#exit

Switch1#

Layer 3 Switch2:

Configure the IP address for interface vlan1 and vlan2.
Switch2#config
Switch2(config)# interface vlan 1
Switch2(config-if-vlan1)# ip address 10.1.1.2 255.255.255.0
Switch2(config-if-vlan1)#no shutdown
Switch2(config-if-vlan1)#exit
Switch2(config)# interface vlan 3
Switch2(config-if-vlan3)# ip address 20.1.1.1 255.255.255.0
Switch2(config-if-vlan3)#no shutdown
Switch2(config-if-vlan3)#exit
Enable OSPF protocol, configure the OSPF area interfaces vlan1 and vlan3 in
Switch2(config)#router ospf
Switch2(config-router)# network 10.1.1.0/24 area 0
Switch2(config-router)# network 20.1.1.0/24 area 1
Switch2(config-router)#exit
Switch2(config)#exit
Switch2#
Layer 3 Switch3:
Configuration of the IP address for interface vlan3.
Switch3#config
Switch3(config)# interface vlan 3
Switch3(config-if-vlan1)# ip address 20.1.1.2 255.255.255.0
Switch3(config-if-vlan3)#no shutdown
Switch3(config-if-vlan3)#exit
Initiate the OSPF protocol, configure the OSPF area to which interface vlan3 belongs
Switch3(config)#router ospf
Switch3(config-router)# network 20.1.1.0/24 area 1
Switch3(config-router)#exit
Switch3(config)#exit
Switch3#
Layer 3 Switch4:
Configuration of the IP address for interface vlan3
Switch4#config
Switch4(config)# interface vlan 3
Switch4(config-if-vlan3)# ip address30.1.1.2 255.255.255.0
Switch4(config-if-vlan3)#no shutdown
Switch4(config-if-vlan3)#exit
Enable OSPF protocol, configure the OSPF area interfaces vlan3 resides in.
Switch4(config)#router ospf
Switch4(config-router)# network 30.1.1.0/24 area 0
Switch4(config-router)#exit
Switch4(config)#exit
Switch4#
Layer 3 Switch5:

Configuration of the IP address for interface vlan2

Switch5#config

Switch5(config)# interface vlan 2

Switch5(config-if-vlan2)# ip address 100.1.1.2 255.255.255.0

Switch5(config-if-vlan2)#no shutdown

Switch5(config-if-vlan2)#exit

Configuration of the IP address for interface vlan3

Switch5(config)# interface vlan 3

Switch5(config-if-vlan3)# ip address 30.1.1.1 255.255.255.0

Switch5(config-if-vlan3)#no shutdown

Switch5(config-if-vlan3)#exit

Enable OSPF protocol, configure the number of the area in which interface vlan2 and vlan3 reside in.

Switch5(config)#router ospf

Switch5(config-router)# network 30.1.1.0/24 area 0

Switch5(config-router)# network 100.1.1.0/24 area 0

Switch5(config-router)#exit

Switch5(config)#exit

Switch5#

Scenario 2: Typical OSPF protocol complex topology.

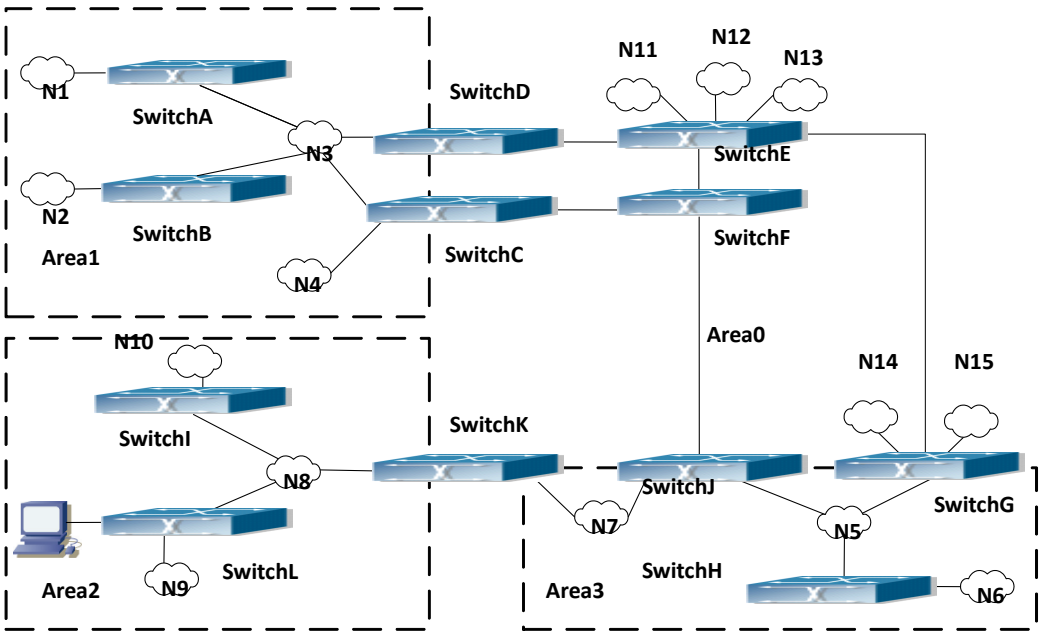


Figure 4-6 Typical complex OSPF autonomous system

This scenario is a typical complex OSPF autonomous system network topology. Area1 include network N1-N4 and layer3 SwitchA-SwitchD, area2 include network N8-N10, host H1 and layer3 SwitchH, area3 include N5-N7 and layer3 SwitchF, SwitchG SwitchA0 and Switch11, and network N8-N10 share a summary route with host H1(i.e. area3 is defined as a STUB area). Layer3 SwitchA,

SwitchB, SwitchD, SwitchE, SwitchG, SwitchH, Switch12 are in-area layer3 switches, SwitchC, SwitchD, SwitchF, Switch10 and Switch11 are edge layer3 switches of the area, SwitchD and SwitchF are edge layer3 switches of the autonomous system.

To area1, layer3 switches SwitchA and SwitchB are both in-area switches, area edge switches SwitchC and SwitchD are responsible for reporting distance cost to all destination outside the area, while they are also responsible for reporting the position of the AS edge layer3 switches SwitchD and SwitchF, AS exterior link-state advertisement from SwitchD and SwitchF are flooded throughout the whole autonomous system. When ASE LSA floods in area 1, those LSAs are included in the area 1 database to get the routes to network N11 and N15.

In addition, layer3 SwitchC and SwitchD must summary the topology of area 1 to the backbone area (area 0, all non-0 areas must be connected via area 0, direct connections are not allowed), and advertise the networks in area 1 (N1-N4) and the costs from SwitchC and SwitchD to those networks. As the backbone area is required to keep connected, there must be a virtual link between backbone layer3 Switch10 and Switch11. The area edge layer3 switches exchange summary information via the backbone layer3 switch, each area edge layer3 switch listens to the summary information from the other edge layer3 switches.

Virtual link can not only maintain the connectivity of the backbone area, but also strengthen the backbone area. For example, if the connection between backbone layer3 SwitchG and Switch10 is cut down, the backbone area will become incontinuous. The backbone area can become more robust by establishing a virtual link between backbone layer3 switches SwitchF and Switch10. In addition, the virtual link between SwitchF and Switch10 provide a short path from area 3 to layer3 SwitchF.

Take area 1 as an example. Assume the IP address of layer3 SwitchA is 10.1.1.1, IP address of layer3 SwitchB interface VLAN2 is 10.1.1.2, IP address of layer3 SwitchC interface VLAN2 is 10.1.1.3, IP address of layer3 SwitchD interface VLAN2 is 10.1.1.4. SwitchA is connecting to network N1 through Ethernet interface VLAN1 (IP address 20.1.1.1); SwitchB is connecting to network N2 through Ethernet interface VLAN1 (IP address 20.1.2.1); SwitchC is connecting to network N4 through Ethernet interface VLAN3 (IP address 20.1.3.1). All the three addresses belong to area 1. SwitchC is connecting to layer3 SwitchE through Ethernet interface VLAN1 (IP address 10.1.5.1); SwitchD is connecting to layer3 SwitchD through Ethernet interface VLAN1 (IP address 10.1.6.1); both two addresses belong to area 1. Simple authentication is implemented among layer3 switches in area1, edge layer3 switches of area 1 authenticate with the area 0 backbone layer3 switches by MD5 authentication.

The followings are just configurations for all layer3 switches in area 1, configurations for layer3 switches of the other areas are omitted. The following are the configurations of SwitchA SwitchB.SwitchC and SwitchD:

1)SwitchA:

Configure IP address for interface vlan2

SwitchA#config

SwitchA(config)# interface vlan 2

SwitchA(config-If-Vlan2)# ip address 10.1.1.1 255.255.255.0

SwitchA(config-If-Vlan2)#exit

Enable OSPF protocol, configure the area number for interface vlan2.

SwitchA(config)#router ospf

SwitchA(config-router)#network 10.1.1.0/24 area 1
SwitchA(config-router)#exit
Configure simple key authentication.
SwitchA(config)#interface vlan 2
SwitchA(config-If-Vlan2)#ip ospf authentication
SwitchA(config-If-Vlan2)#ip ospf authentication-key test
SwitchA(config-If-Vlan2)exit
Configure IP address and area number for interface vlan1.
SwitchA(config)# interface vlan 1
SwitchA(config-If-Vlan1)#ip address 20.1.1.1 255.255.255.0
SwitchA(config-If-Vlan1)#exit
SwitchA(config)#router ospf
SwitchA(config-router)#network 20.1.1.0/24 area 1
SwitchA(config-router)#exit
2)SwitchB:
Configure IP address for interface vlan2
SwitchB#config
SwitchB(config)# interface vlan 2
SwitchB(config-If-Vlan2)# ip address 10.1.1.2 255.255.255.0
SwitchB(config-If-Vlan2)#exit
Enable OSPF protocol, configure the area number for interface vlan2.
SwitchB(config)#router ospf
SwitchB(config-router)#network 10.1.1.0/24 area 1
SwitchB(config-router)#exit
SwitchB(config)#interface vlan 2
Configure simple key authentication.
SwitchB(config)#interface vlan 2
SwitchB(config-If-Vlan2)#ip ospf authentication
SwitchB(config-If-Vlan2)#ip ospf authentication-key test
SwitchB(config-If-Vlan2)#exit
Configure IP address and area number for interface vlan1.
SwitchB(config)# interface vlan 1
SwitchB(config-If-Vlan1)#ip address 20.1.2.1 255.255.255.0
SwitchB(config-If-Vlan1)#exit
SwitchB(config)#router ospf
SwitchB(config-router)#network 20.1.2.0/24 area 1
SwitchB(config-router)#exit
SwitchB(config)#exit
3)SwitchC:
Configure IP address for interface vlan2
SwitchC#config
SwitchC(config)# interface vlan 2
SwitchC(config-If-Vlan2)# ip address 10.1.1.3 255.255.255.0
SwitchC(config-If-Vlan2)#exit

Enable OSPF protocol, configure the area number for interface vlan2
SwitchC(config)#router ospf
SwitchC(config-router)#network 10.1.1.0/24 area 1
SwitchC(config-router)#exit
Configure simple key authentication
SwitchC(config)#interface vlan 2
SwitchC(config-If-Vlan2)#ip ospf authentication
SwitchC(config-If-Vlan2)#ip ospf authentication-key test
SwitchC(config-If-Vlan2)#exit
Configure IP address and area number for interface vlan3
SwitchC(config)# interface vlan 3
SwitchC(config-If-Vlan3)#ip address 20.1.3.1 255.255.255.0
SwitchC(config-If-Vlan3)#exit
SwitchC(config)#router ospf
SwitchC(config-router)#network 20.1.3.0/24 area 1
SwitchC(config-router)#exit
Configure IP address and area number for interface vlan 1
SwitchC(config)# interface vlan 1
SwitchC(config-If-Vlan1)#ip address 10.1.5.1 255.255.255.0
SwitchC(config-If-Vlan1)#exit
SwitchC(config)#router ospf
SwitchC(config-router)#network 10.1.5.0/24 area 0
SwitchC(config-router)#exit
Configure MD5 key authentication.
SwitchC(config)#interface vlan 1
SwitchC (config-If-Vlan1)#ip ospf authentication message-digest
SwitchC (config-If-Vlan1)#ip ospf authentication-key test
SwitchC (config-If-Vlan1)#exit
SwitchC(config)#exit
SwitchC#
4)SwitchD:
Configure IP address for interface vlan2
SwitchD#config
SwitchD(config)# interface vlan 2
SwitchD(config-If-Vlan2)# ip address 10.1.1.4 255.255.255.0
SwitchD(config-If-Vlan2)#exit
Enable OSPF protocol, configure the area number for interface vlan2.
SwitchD(config)#router ospf
SwitchD(config-router)#network 10.1.1.0/24 area 1
SwitchD(config-router)#exit
Configure simple key authentication.
SwitchD(config)#interface vlan 2
SwitchD(config-If-Vlan2)#ip ospf authentication
SwitchD(config-If-Vlan2)#ip ospf authentication-key test

SwitchD(config-If-Vlan2)#exit
Configure the IP address and the area number for the interface vlan 1
SwitchD(config)# interface vlan 1
SwitchD(config-If-Vlan1)# ip address 10.1.6.1 255.255.255.0
SwitchD(config-If-Vlan1)exit
SwitchD(config)#router ospf
SwitchD(config-router)#network 10.1.6.0/24 area 0
SwitchD(config-router)#exit
Configure MD5 key authentication
SwitchD(config)#interface vlan 1
SwitchD(config-If-Vlan1)#ip ospf authentication message-digest
SwitchD(config-If-Vlan1)#ip ospf authentication-key test
SwitchD(config-If-Vlan1)exit
SwitchD(config)#exit
SwitchD#

Scenario 3: The function of OSPF importing the routers of other OSPF processes

As shown in the following graph, a switch running the OSPF routing protocol connects two networks: network A and network B. Because of some reason, it is required that network A should be able to learn the routers of network B, but network B should not be able to learn the routers of network A. According to that, two OSPF processes can be started respectively on interface vlan 1 and interface vlan 2. the OSPF process which interface vlan 1 belongs to is configured to import the routers of the OSPF process which interface vlan 2 belongs to, while the OSPF process which interface vlan 2 belongs to should not be configured to import the routers of the OSPF process which interface vlan 1 belongs to.

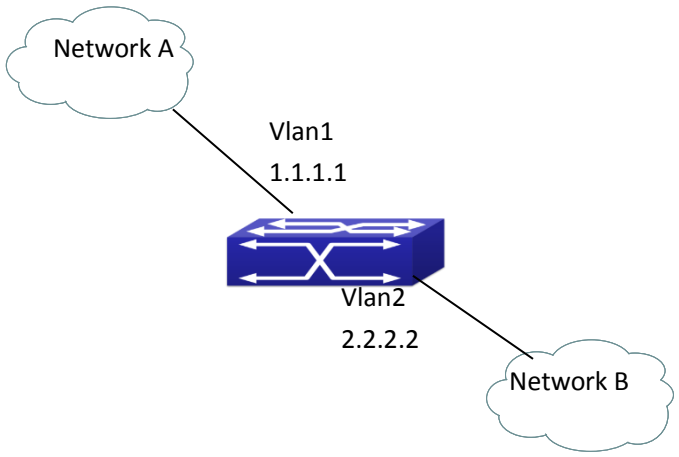


Figure 4-7 Function of OSPF importing the routers of other OSPF processes example

We can configure as follows:
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 1.1.1.1 255.255.255.0
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 2

Switch(Config-if-Vlan2)#ip address 2.2.2.2 255.255.255.0
Switch(Config-if-Vlan2)#exit
Switch(config)#router ospf 10
Switch(config-router)#network 2.2.2.0/24 area 1
Switch(config-router)#exit
Switch(config)#router ospf 20
Switch(config-router)#network 1.1.1.0/24 area 1
Switch(config-router)#redistribute ospf 10
Switch(config-router)#exit

4.4.3.2 Configuration Examples of OSPF VPN

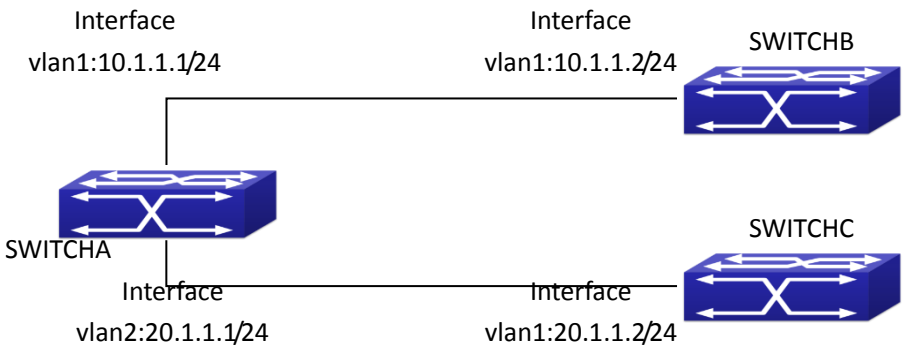


Figure 4-8 OSPF VPN Example

The above figure shows that a network consists of three Layer 3 switches in which the switchA as PE, SwitchB and SwitchC as CE1 and CE2. The PE is connected to CE1 and CE2 through vlan1 and vlan2. The routing messages are exchanged between PE and CE through OSPF protocol.

a) SwitchA, the Layer 3 switch as PE

Configure VPN route/transmitting examples vpnb and vpnc
SwitchA#config
SwitchA(config)#ip vrf vpnb
SwitchA(config-vrf)#
SwitchA(config-vrf)#exit
SwitchA#(config)
SwitchA(config)#ip vrf vpnc
SwitchA(config-vrf)#
SwitchA(config-vrf)#exit
Associate the vlan 1 and vlan 2 respectively with vpnb and vpnc while configuring IP address
SwitchA(config)#in vlan1
SwitchA(config-if-Vlan1)#ip vrf forwarding vpnb
SwitchA(config-if-Vlan1)#ip address 10.1.1.1 255.255.255.0
SwitchA(config-if-Vlan1)#exit
SwitchA(config)#in vlan2
SwitchA(config-if-Vlan2)#ip vrf forwarding vpnc
SwitchA(config-if-Vlan2)#ip address 20.1.1.1 255.255.255.0

```
SwitchA(config-if-Vlan2)#exit
Configure OSPF examples associated with vpnb and vpnc respectively
SwitchA(config)#
SwitchA(config)#router ospf 100 vpnb
SwitchA(config-router)#network 10.1.1.0/24 area 0
SwitchA(config-router)#redistribute bgp
SwitchA(config-router)#exit
SwitchA(config)#router ospf 200 vpnc
SwitchA(config-router)#network 20.1.1.0/24 area 0
SwitchA(config-router)#redistribute bgp
```

b) The Layer 3 SwitchB of CE1：

```
Configure the IP address of Ethernet E 1/0/2
SwitchB#config
SwitchB(config)# interface Vlan1
SwitchB(config-if-vlan1)# ip address 10.1.1.2 255.255.255.0
SwitchB (config-if-vlan1)exit
Enable OSPF protocol and configuring OSPF segments
SwitchB(config)#router ospf
SwitchB(config-router-rip)#network 10.1.1.0/24 area 0
SwitchB(config-router-rip)#exit
```

c) The Layer 3 SwitchC of CE2

```
Configure the IP address of Ethernet E 1/0/2
SwitchC#config
SwitchC(config)# interface Vlan1
SwitchC(config-if-vlan1)# ip address 20.1.1.2 255.255.255.0
SwitchC(config-if-vlan1)#exit
Initiate OSPF protocol and configuring OSPF segments
SwitchC(config)#router ospf
SwitchC(config-router)#network 20.1.1.0/24 area 0
SwitchC(config-router)#exit
```

4.4.4 OSPF Troubleshooting

The OSPF protocol may not be working properly due to errors such as physic connection, configuration error when configuring and using the OSPF protocol. So users should pay attention to following:

- ☞ First ensure the physic connection is correct
- ☞ Second, ensure the interface and link protocol are UP (use **show interface** command)
- ☞ Configure different IP address from different segment on each interface
- ☞ Then initiate OSPF protocol (use **router-ospf** command) and configure the OSPF area on corresponding interface
- ☞ After that, a OSPF protocol feature should be checked---the OSPF backbone area should be

continuous and apply virtual link to ensure it is continuous. if not; all non 0 areas should only be connected to other non 0 area through 0 area; a border Layer 3 switch means that one part of the interfaces of this switch belongs to 0 area, the other part belongs to non 0 area; Layer 3 switch DR should be specified for multi-access network such as broadcast network.

4.5 BGP

4.5.1 Introduction to BGP

BGP stands for a Border Gateway Protocol. It's a dynamic routing protocol inter-autonomous system. Its basic function is automatically exchanging routing information without loops. By exchanging routing reachable information with autonomous number of AS sequence attributes, BGP could create autonomous topological map to eliminate routing loop and implement policies configured by users. Generally, the switches in an AS may use several IGPs (Interior Gateway Protocol) in order to exchange routing information in the AS, such as RIP and OSPF which are IGPs; and exchange information among ASes with EGP (Exterior Gateway Protocol). For example, BGP is one kind of EGP. The AS is usually established on a single administrative department. BGP is often used on the switches among ISPs or the departments of Multi-national Corporation.

BGP has been used since1989, its earliest three versions are RFC1105 (BGP-1), RFC1163 (BGP-2) and RFC1267 (BGP-3).Currently, the most popular one is RFC1771 (BGP-4). The switch supports BGP-4.

1. Characteristics of BGP-4

BGP-4 is suitable for the distributed structure and supports Classless InterDomain Routing (CIDR). BGP-4 is becoming the virtual exterior routing protocol standard used for the global Internet. The features of BGP-4 are as follows.

- BGP is an exterior routing protocol, unlike interior routing protocol, such as OSPF and RIP, BGP can't discovery and calculate routes, but it can control the transmission of routes and select the best route.
- By carrying AS routing information in the updating route, the problem of Routing Loops can be resolved
- BGP uses TCP on port 179 as its transport protocol, this could enhance the reliability of the protocol.
- BGP-4 supports CIDR (Classless InterDomain Routing), which is an important improvement to BGP-3. CIDR has a brand new way to look on IP address; it doesn't distinguish class A , Class B and class C network. For instance, an illegal class C address 192.213.0.0 255.255.0.0 can be represented as 192.213.0.0/16 by CIDR which is a legal super network. /16 represents that the network number is formed by 16 bits from the

beginning left of the address. The introduction of CIDR abbreviates the route aggregation. The route aggregation is the process of combining several different routes. So notifying several routes can be changed to notify only one route which decreases the route table.

- When updating route, BGP send only incremental route. The bandwidth occupied by BGP transmission is reduced greatly and it is suitable for the mass routing information transmitted on the internet
- For political and economical reasons, each AS expects to filter and control the route, BGP-4 provides abundant route policies which make BGP-4 more extendable to encourage the internet development.

2. The Overview of BGP-4 operation

Unlike RIP and OSPF protocols, BGP protocol is connection oriented. BGP switches must establish connection to exchange routing information. The operation of BGP protocol is driven by messages and the messages can be divided into four kinds:

Open message----It's the first message which is sent after a TCP connection is established. It is used to create BGP connecting relation among BGP peers. Some parameters in Open Message are used to negotiate if a connection could be established among BGP peers.

Keepalive Message ----- it's the message to check connection availability. It's usually sent periodically to keep BGP connection. If this message or Update message is not received within holdtime time, BGP connection is closed.

Update Message----- it's the most important message in the BGP system. It's used to exchange routing information among peers. The switches exchange not only updated routing information, but also unavailable or canceled routing information. It consists of three parts: unreachable route, NLRI (Network LayerReachability Information) and Path Attributes.

Notification Message------it's the mistake notification message. When a BGP speaker receives this message, it shutdowns the BGP connections with its neighbors

BGP-4 is connection oriented. BGP acts as higher protocol and runs on the particular equipments. When detecting a neighbor, a TCP session is established and maintained. Then the exchanging and synchronization of the route table will be carried out. By sending the whole BGP route table the routing information is exchanged only when the system initiates. After that, the routing information is exchanged only when the updated routing information is available. Only incremental update message is exchanged. BGP-4 maintains links and sessions periodically through keep alive message. That is sending and receiving keep alive message periodically to check if the connections are normal.

The switches that participate the BGP session are called BGP speaker. It continuously receives or generates new routing information and advertises it to other BGP speakers. When a BGP speaker receives a new routing notification from other AS, if this route is better than the presently known route or there is no acceptable route, it sends this route to all the other BGP speakers of the AS. A BGP speaker calls other speakers that exchange route information with it as neighbors or peers. Several relevant neighbors can constitute a peer group. BGP operates on the switches in the following two manners:

- IBGP：Internal BGP

- EBGP：External BGP

When BGP runs in the same AS, it's called IBGP. When in the different AS, it's called EBGP. Generally, the outer neighbors are connected physically and the inner neighbors can be in any place of the AS. The difference is finally shown in the dealing manner of BGP to routing information. The equipments may check the AS numbers of the Open Message from neighbors to decide treating the neighbor switches as the exterior neighbor or as the interior neighbor.

IBGP are used in the AS. It sends message to all the BGP neighbors in the AS. IBGP exchanges AS routing information in a big organization. Attention, the switches in the AS needn't be connected physically. Only if the switches are in the same AS, they can be neighbors each other. Because BGP can't detect route, the route tables of other inner route protocols (such as static route, direct route, OSPF and RIP) need contain neighbor IP addresses and these routes are used to exchange information among BGPs. In order to avoid routing loops, when a BGP speaker receives a route notification from inner neighbor, it would not notify this route to other inner neighbors.

EBGP is used among the AS, and it transmits routing information to the BGP neighbors of outer ASes. EBGP need physical connection and share the same medium. Because EBGP need physical connection, the boundary equipments between two AS are usually running EBGP. When a BGP speaker receives routing information from outer neighbors, it notifies these routes to other inner neighbors.

3． Route attribute

BGP-4 can share and query inner IP route table through relevant mechanisms, but it has its own route table. In the BGP route table, each route has a network number, AS listing information (also called AS path) that it passed and some routing attributes (such as origin). The routing attribute that BGP-4 used is very complex, this attribute can be used as metrics to select path.

4． Route-selecting policy of BGP

When receiving BGP notification about a same route from several neighbors, selecting the best route need to be take into account after routing filtering. This process is called BGP route selecting process. BGP route selecting process will start only when the following conditions are fulfilled:

- The switch's route must be next hop reachable. That is in the route table there is the route that can reach the next hop.
- BGP must be synchronized with IGP (unless asynchronism is configured; only restricted to IBGP)

BGP route selecting process is based on the BGP attribute. When there are several routes that indicate the same destination, BGP need select the best route to the destination. The decision-making process is as the following:

1．Select the route with the most weight first;

2. If the weights are the same, select the route with the most local preference;

3. If the local preferences are the same, select the route generated by local switch.

4. If the local preferences are the same and there is no route generated by local switch, select the route with the shortest AS path;

5. If the AS paths are the same, select the route with the lowest "origin" type (IGP<EGP<INCOMPLETE);

6. If the “origin” types are the same, select the route with the lowest MED attribute. Unless activating command “bgp always-compare-med”, this comparison is only available among the routes from the same neighbor AS.

7. If the MED attributes are the same, EBGP is preferable to outer confederation and outer confederation is preferable to IBGP.

8. If it’s still the same by now, BGP router ID （router ID）is used to break the balance. The best route is the one from the least router ID.

9. If it’s still the same by now, BGP router ID （router ID）is used to break the balance. The best route is the one from the least router ID.

4.5.2 BGP Configuration Task List

The BGP configuration tasks include basic and advanced tasks. Basic BGP configuration tasks include the following:

1. Enable BGP Routing (required)
2. Configure BGP Neighbors (required)
3. Administrate the change of routing policy
4. Configure BGP Weights
5. Configure BGP Route Filtering policy basing on Neighbors
6. Configure Next-Hop of BGP
7. Configure Multi-Hop of EGBP
8. Configure BGP Session Identifier
9. Configure BGP Version

Advanced BGP configuration tasks include the following:

1. Use Route Maps to Modify Route
2. Configure Route Aggregation
3. Configure BGP Community Filtering
4. Configure BGP Confederation
5. Configure a Route Reflector
6. Configure Peer Groups
7. Configure Neighbors and Peer Groups’ Parameters
8. Adjust BGP Timers
9. Adjust BGP Announcement Interval
10. Configure the default Local Priority
11. Allow to Transfer Default Route
12. Configure BGP’s MED Value
13. Configure BGP Routing Redistribution
14. Configure BGP Route Dampening
15. Configure BGP capability Negotiation

16. Configure Routing Server
17. Configure Path-Selected Rule
18. Configure redistribution of OSPF routing to BGP
    (1) Enable redistribution of OSPF routing to BGP
    (2) Display and debug the information about configuration of redistribution of OSPF routing to BGP

Ⅰ. Basic BGP configuration tasks

1. Enable BGP Routing

Command	Explanation
Global mode	
router bgp <*as-id*> **no router bgp <*as-id*>**	Enable BGP, the **no router bgp <*as-id*>** command disables BGP process.
BGP protocol mode	
bgp *asnotation asdot* ***no* bgp asnotation asdot**	Show AS number and match the regular expression **with ASDOT method*. The no*** command cancels this method.
network <ip-address/M> **no network <ip-address/M>**	Set the network that BGP will announce, the no network <ip-address/M> command cancels the network that will be announced.
address-family ipv4 {unicast| multicast|vrf <*vrf-nam*>} **no address-family ipv4 {unicast| multicast|vrf <*vrf-nam*>}**	Create IPv4 for BGP protocol and enter BGP-VPN view. Any IPv4 is not created by default.

2. Configure BGP Neighbors

Command	Explanation
Router configuration mode	
neighbor {<*ip-address*>|<*TAG*>} remote-as <*as-id*> **no neighbor {<*ip-address*>|<*TAG*>} [remote-as <*as-id*>]**	Specify a BGP neighbor, the **no neighbor {<*ip-address*>|<*TAG*>} [remote-as <*as-id*>]** command deletes the neighbor.

3. Administrate the change of routing policy

（1）Configure hard reconfiguration.

Command	Explanation
Admin Mode	
clear ip bgp {<*>|<*as-id*>| external|peer-group <*NAME*>|<*ip-address*>}	Configure hard reconfiguration.

（2） Configure outbound soft reconfiguration.

Command	Explanation
Admin Mode	
clear ip bgp {<*>|<*as-id*>| external|peer-group <NAME>|<*ip-address*>} soft out	Configure outbound soft reconfiguration.

（3）Configure inbound soft reconfiguration.

Command	Explanation
BGP configuration mode	
neighbor { *<ip-address>* | *<TAG>* } soft-reconfiguration inbound **no neighbor { *<ip-address>* | *<TAG>* } soft-reconfiguration inbound**	This command can store routing information from neighbors and peers; the **no neighbor { *<ip-address>* | *<TAG>* } soft-reconfiguration inbound** command cancels the storage of routing information.
Admin Mode	
clear ip bgp {*<*>*|*<as-id>*| external|peer-group *<NAME>*|*<ip-address>*} soft in	Configure BGP inbound soft reconfiguration.

4. Configure BGP Weights

Command	Explanation
BGP configuration mode	
neighbor { *<ip-address>* | *<TAG>* } weight *<weight>* **no neighbor { <ip-address> | <TAG> }**	Configure BGP neighbor weights; the **no neighbor { <ip-address> | <TAG> }** command recovers default weights.

5. Configure BGP Route Filtering policy based on neighbor

Command	Explanation
BGP configuration mode	
neighbor {*<ip-address>*|*<TAG>*} distribute-list {*<1-199>*|*<1300-2699>*|*<WORD>*} {in|out} **no neighbor {*<ip-address>*|*<TAG>*} distribute-list {*<1-199>*|*<1300-2699>*|*<WORD>*} {in|out}**	Filter neighbor routing updating information. The **no neighbor {*<ip-address>* | *<TAG>*} distribute-list {*<1-199>*|*<1300-2699>*| *<WORD>*} {in|out}** command cancels routing filter.

6. Configure Next-Hop

1) Set Next-Hop as the switch's address

Command	Explanation
BGP configuration mode	

neighbor { *<ip-address>* / *<TAG>* } next-hop-self **no neighbor { *<ip-address>* / *<TAG>* } next-hop-self**	While sending route Next-Hop set Next-Hop as the switch's address; the **no neighbor { *<ip-address>* / *<TAG>* } next-hop-self** command cancels the setting.

2）Cancel default Next-Hop through route map

Command	Explanation
Route mapped configuration command	
set ip next-hop *<ip-address>* **no set ip next-hop**	Set the Next-Hop attribute of outbound route. The **no set ip next-hop** command cancels this setting.

7. Configure EGBP Multi-Hop

If the connections with outer neighbors are not direct, the following command can configure neighbor Multi-Hop.

Command	Explanation
BGP configuration mode	
neighbor {*<ip-address>*|*<TAG>*} ebgp-multihop [*<1-255>*] **no neighbor {*<ip-address>*|*<TAG>*} ebgp-multihop [*<1-255>*]**	Configure the allowance of EBGP connection with other networks that are not connected directly; the **no neighbor {*<ip-address>*|*<TAG>*} ebgp-multihop [*<1-255>*]** command cancels the setting.

8. Configure BGP session identifier

Command	Explanation
BGP configuration mode	
bgp router-id *<ip-address>* **no bgp router-id**	Configure the router-id value; the **no bgp router-id** command recovers the default value.

9. Configure the BGP Version

Command	Explanation
BGP configuration mode	
neighbor {*<ip-address>* / *<TAG>*} version <value> **no neighbor {*<ip-address>* / *<TAG>*} version**	Set the version used by BGP neighbors; the **no neighbor {*<ip-address>* / *<TAG>*} version** command recovers default setting. Presently only supporting version 4^{th}.

Ⅱ. Advanced BGP configuration tasks

1. Use Route Maps to Modify Route

Command	Explanation
BGP configuration mode	
neighbor { *<ip-address>* / *<TAG>* } route-map *<map-name >* {in | out} **no neighbor { *<ip-address>* / *<TAG>* } route-map *<map-name >* {in | out}**	Apply a route map to incoming or outgoing routes; the **no neighbor { *<ip-address>* / *<TAG>* } route-map *<map-name >* {in | out}** command cancels the settings of routing maps.

2. Configure Route Aggregation

Command	Explanation
BGP configuration mode	
aggregate-address *<ip-address/M>* [summary-only] [as-set] **no aggregate-address *<ip-address/M>* [summary-only] [as-set]**	Create an aggregate entry in the BGP routing table; the **no aggregate-address *<ip-address/M>* [summary-only] [as-set]** command cancels the aggregate entry.

3. Configure BGP Community Filtering

Command	Explanation
BGP configuration mode	
neighbor {*<ip-address>* / *<TAG>*} send-community **no neighbor {*<ip-address>* / *<TAG>*} send-community**	Allow the routing updates with community attributes sending to BGP neighbors; the **no neighbor {*<ip-address>* / *<TAG>*} send-community** command enables the route without community attributes.

4. Configure BGP Confederation

Command	Explanation
BGP configuration mode	
bgp confederation identifier *<as-id>* **no bgp confederation identifier *<as-id>***	Configure a BGP AS confederation identifier; the **no bgp confederation identifier *<as-id>*** command deletes the BGP AS confederation identifier.
bgp confederation peers *<as-id>* [*<as-id>..*] **no bgp confederation peers *<as-id>* [*<as-id>..*]**	Configure the AS affiliated to the AS confederation; the **no bgp confederation peers *<as-id>* [*<as-id>..*]** command deletes the AS from the AS confederation.

5. Configure a Route Reflector

（1） The following commands can be used to configure route reflector and its clients.

Command	Explanation
BGP configuration mode	
neighbor *<ip-address>* route-reflector-client **no neighbor *<ip-address>* route-reflector-client**	Configure the current switch as route reflector and specify a client; the **no neighbor *<ip-address>* route-reflector-client** command format deletes a client.

（2） If there are more than one route reflectors in the cluster, the following commands can configure cluster-id

Command	Explanation
BGP configuration mode	
bgp cluster-id *<cluster-id>* **no bgp cluster-id**	Configure cluster id; the **no bgp cluster-id** command cancels the cluster id configuration.

（3） If the route reflector from clients to clients is needed, the following commands can be used.

Command	Explanation
BGP configuration mode	
bgp client-to-client reflection **no bgp client-to-client reflection**	Configure the allowance of the route reflector from clients to clients; the **no bgp client-to-client reflection** command forbids this allowance.

6．Configure Peer Groups

（1） Create peer groups

Command	Explanation
BGP configuration mode	
neighbor *<TAG>* peer-group **no neighbor *<TAG>* peer-group**	Create peer groups; the **no neighbor *<TAG>* peer-group** command deletes peer groups.

（2） Add neighbors to peers groups

Command	Explanation
BGP configuration mode	
neighbor *<ip-address>* peer-group *<TAG>* **no neighbor *<ip-address>* peer-group *<TAG>***	Make a neighbor a member of the peer group. The **no neighbor *<ip-address>* peer-group *<TAG>*** command cancels the specified member.

7．Configure neighbors and peer Groups' parameters

Command	Explanation
BGP configuration mode	
neighbor {*<ip-address>* / *<TAG>*} remote-as *<as-id>* **no neighbor {*<ip-address>* / *<TAG>*} remote-as *<as-id>***	Specify a BGP neighbor; the **no neighbor {*<ip-address>* / *<TAG>*} remote-as *<as-id>*** command deletes the neighbor.
neighbor { *<ip-address>* / *<TAG>* } description *<.LINE>* **no neighbor { *<ip-address>* / *<TAG>* } description**	Associate a description with a neighbor; the **no neighbor {*<ip-address>* / *<TAG>*} description** command deletes this description.
neighbor { *<ip-address>* / *<TAG>* } default-originate [route-map *<NAME>*] **no neighbor { *<ip-address>* / *<TAG>* } default-originate [route-map *<NAME>*]**	Permit to send the default route 0.0.0.0; the **no neighbor { *<ip-address>* / *<TAG>* } default-originate [route-map *<NAME>*]** command cancels sending default route.
neighbor { *<ip-address>* / *<TAG>* } send-community **no neighbor { *<ip-address>* / *<TAG>* } send-community**	Configure the community attributes sent to the neighbor.
neighbor { *<ip-address>* / *<TAG>* } timers *<keep alive> <holdtime>* **no neighbor { *<ip-address>* / *<TAG>* } timers**	Configure a particular neighbor's keep-alive and hold-time timer; the **no neighbor {*<ip-address>* / *<TAG>*} timers** command recovers the default value.
neighbor {*<ip-address>* / *<TAG>*} advertisement-interval *<seconds>* **no neighbor {*<ip-address>* / *<TAG>*} advertisement-interval**	Configure the min interval of sending BGP routing information; the **no neighbor {*<ip-address>* / *<TAG>*} advertisement-interval** command recovers the default value.
neighbor {*<ip-address>* / *<TAG>*} ebgp-multihop [*<1-255>*] **no neighbor {*<ip-address>* / *<TAG>*} ebgp-multihop**	Configure the allowance of EBGP connections with networks connected indirectly; the **no neighbor {*<ip-address>* / *<TAG>*} ebgp-multihop** command cancels this setting.
neighbor { *<ip-address>* / *<TAG>* } weight <weight> **no neighbor { *<ip-address>* / *<TAG>* } weight**	Configure BGP neighbor weights; the **no neighbor { *<ip-address>* / *<TAG>* } weight** command recovers the default weights.
**neighbor { *<ip-address>* / *<TAG>* } distribute-list { *<access-list-number>*	*<name>* } { in	out }** **no neighbor { *<ip-address>* / *<TAG>* } distribute-list { *<access-list-number>*	*<name>* } { *in*	*out* }**	Filter neighbor route update; the **no neighbor { *<ip-address>* / *<TAG>* } distribute-list { *<access-list-number>*	*<name>* } { *in*	*out* }** command cancels route filtering.

neighbor { *<ip-address>* / *<TAG>* } route-reflector-client **no neighbor { *<ip-address>* / *<TAG>* } route-reflector-client**	Configure the current switch as route reflector and specify a client; the **no neighbor { *<ip-address>* / *<TAG>* } route-reflector-client** command deletes a client.
neighbor { *<ip-address>* / *<TAG>* } next-hop-self **no neighbor { *<ip-address>* / *<TAG>* } next-hop-self**	When sending route, configure Next-Hop as its address; the **no neighbor { *<ip-address>* / *<TAG>* } next-hop-self** command cancels the setting.
neighbor { *<ip-address>* / *<TAG>* } version *<value>* **no neighbor { *<ip-address>* / *<TAG>* } version**	Specify the BGP version communicating with BGP neighbors; the **no neighbor { *<ip-address>* / *<TAG>* } version** command recovers default setting.
neighbor { *<ip-address>* / *<TAG>* } route-map *<map-name>* {in | out} **no neighbor { *<ip-address>* / *<TAG>* } route-map *<map-name>* {in | out}**	Apply a route map to incoming or outgoing routes; the **no neighbor { *<ip-address>* / *<TAG>* } route-map *<map-name>* {in | out}** command cancels the setting of route reflector.
neighbor { *<ip-address>* / *<TAG>* } soft-reconfiguration inbound **no neighbor { *<ip-address>* / *<TAG>* } soft-reconfiguration inbound**	Store the route information from neighbor or peers; the **no neighbor { *<ip-address>* / *<TAG>* } soft-reconfiguration inbound** command cancels the storage.
neighbor { *<ip-address>* / *<TAG>* } shutdown **no neighbor { *<ip-address>* / *<TAG>* } shutdown**	Shutdown BGP neighbor or peers; the **no neighbor { *<ip-address>* / *<TAG>* } shutdown** command activates the closed BGP neighbor or peers.

8．Adjust BGP Timers

（1） Configure the BGP timer of all the neighbors

Command	Explanation
BGP configuration mode	
timers bgp *<keep alive>* *<holdtime>* **no timers bgp**	Configure the BGP timers of all the neighbors; the no timer bgp command recovers the default value.

（2） Configure the timer value of a particular neighbor

Command	Explanation
BGP configuration mode	

neighbor { *<ip-address>* / *<TAG>* } timers *<keep alive>* *<holdtime>* **no neighbor { *<ip-address>* / *<TAG>* } timers**	Configure the keep alive and holdtime timer of a particular neighbor; the **no neighbor { *<ip-address>* / *<TAG>* } timers** command recovers the default value.

9. Adjust BGP announcement Interval

Command	Explanation
BGP configuration mode	
neighbor {*<ip-address>* / *<TAG>*} advertisement-interval <seconds> **no neighbor {*<ip-address>* / *<TAG>*} advertisement-interval**	Configure the minimum interval among BGP routes update information; the **no neighbor {*<ip-address>* / *<TAG>*} advertisement-interval** command recovers the default setting.

10. Configure the Local Preference Value

Command	Explanation
BGP configuration mode	
bgp default local-preference *<value>* **no bgp default local-preference**	Change default local preference; the **no bgp default local-preference** command recovers the default value.

11. Enable sending default route

Command	Explanation
BGP configuration mode	
neighbor { *<ip-address>* / *<TAG>* } default-originate **no neighbor { *<ip-address>* / *<TAG>* } default-originate**	Permit sending default route 0.0.0.0; the **no neighbor { *<ip-address>* / *<TAG>* } default-originate** command cancels sending default route.

12. Configure BGP's MED Value

（1） Configure MED value

Command	Explanation
Route map configuration command	
set metric *<metric-value>* **no set metric**	Configure metric value; the **no set metric** command recovers the default value.

（2） Apply route selection based on MED according to the path from different AS

Command	Explanation
BGP configuration mode	

bgp always-compare-med **no bgp always-compare-med**	Permit the MED comparison from different AS; the **no bgp always-compare-med** command forbids the comparison.

13. Configure BGP routing redistribution

Command	Explanation
BGP configuration mode	
redistribute { connected | static | rip | ospf} [metric *<metric>*] [route-map *<NAME>*] **no redistribute { connected | static | rip | ospf}**	Redistribute IGP routes to BGP and may specify the redistributed metric and route reflector; the **no redistribute { connected | static | rip | ospf}** command cancels the redistribution.

14. Configure Route Dampening

Command	Explanation
BGP configuration mode	
bgp dampening [*<1-45>*] [*<1-20000>* *<1-20000>* *<1-255>*] [*<1-45>*] **no bgp dampening**	Enable BGP route dampening and apply the specified parameters; the **no bgp dampening** command stops route dampening

15. Configure BGP capability Negotiation

Command	Explanation
BGP configuration mode	
neighbor {*<ip-address>*|*<TAG>*} capability {dynamic | route-refresh} **no neighbor {*<ip-address>*|*<TAG>*} capability {dynamic | route-refresh}** **neighbor {*<ip-address>*|*<TAG>*} capability orf prefix-list {*<both>*|*<send>*|*<receive>*}** **no neighbor {*<ip-address>*|*<TAG>*} capability orf prefix-list {*<both>*|*<send>*|*<receive>*}** **neighbor {*<ip-address>*|*<TAG>*} dont-capability-negotiate** **no neighbor {*<ip-address>*|*<TAG>*} dont-capability-negotiate** **neighbor {*<ip-address>*|*<TAG>*} override-capability** **no neighbor {*<ip-address>*|*<TAG>*} override-capability** **neighbor {*<ip-address>*|*<TAG>*} strict-capability-match** **no neighbor {*<ip-address>*|*<TAG>*} strict-capability-match**	BGP provides capability negotiation regulation and carry out this capability match while establishing connection. The currently supported capabilities include route update, dynamic capability, outgoing route filtering capability and the address family's capability of supporting the negotiation. Use these command to enable these capabilities, its format "no" close these capabilities .It can also be configured by commands to not do capability negotiation, do strict capability negotiation or not care about the negotiation results.

16. Configure Routing Server

Command	Explanation
BGP configuration mode	
neighbor {*<ip-address>*|*<TAG>*} route-server-client **no neighbor {*<ip-address>*|*<TAG>*} route-server-client**	Route server may configure BGP neighbors under EBGP environment to reduce the number of peers that every client has configured; format "no" of the command configures this router as route server and specify the clients it serves, the **no neighbor {*<ip-address>*|*<TAG>*} route-server-client** command can delete clients.

17. Configure Path-selected rules

Command	Explanation
BGP configuration mode	
bgp always-compare-med **no bgp always-compare-med** **bgp bestpath as-path ignore** **no bgp bestpath as-path ignore** **bgp bestpath compare-confed-aspath** **no bgp bestpath compare-confed-aspath** **bgp bestpath compare-routerid** **no bgp bestpath compare-routerid** **bgp bestpath med {[confed] [missing-is-worst]}** **no bgp bestpath med {[confed] [missing-is-worst]}**	BGP may change some path-select rules by configuration to change the best selection and compare MED under EBGP environment through these command, ignore the AS-PATH length, compare the confederation as-path length, compare the route identifier and compare the confederation MED etc. Its format "no" recovers the default route path-selected rules.

18. Configure redistribution of OSPF routing to BGP

(1) Enable redistribution of OSPF routing to BGP

Command	Explanation
Router BGP Configuration Mode	
redistribute ospf [*<process-id>*] [route-map<*word*>] **no redistribute ospf [<process-id>]**	To enable or disable the redistribution of OSPF routing to BGP.

(2)Display and debug the information about configuration of redistribution of OSPF routing to BGP

Command	Explanation
Admin Mode and Configuration Mode	
show ip bgp redistribute	To enable or disable the redistribution of OSPF routing to BGP.
Admin Mode	

debug bgp redistribute message send **no debug bgp redistribute message send** **debug bgp redistribute route receive** **no debug bgp redistribute route receive**	To enable or disable debugging messages sent by BGP for redistributing OSPF routing. To enable or disable debugging messages received from NSM for redistributing OSPF routing.

4.5.3 Configuration Examples of BGP

4.5.3.1 Examples 1: configure BGP neighbor

SwitchB, SwitchC and SwitchD are in AS200, SwitchA is in AS100. SwitchA and SwitchB share the same network segment. SwitchB and SwitchD are not connected physically.

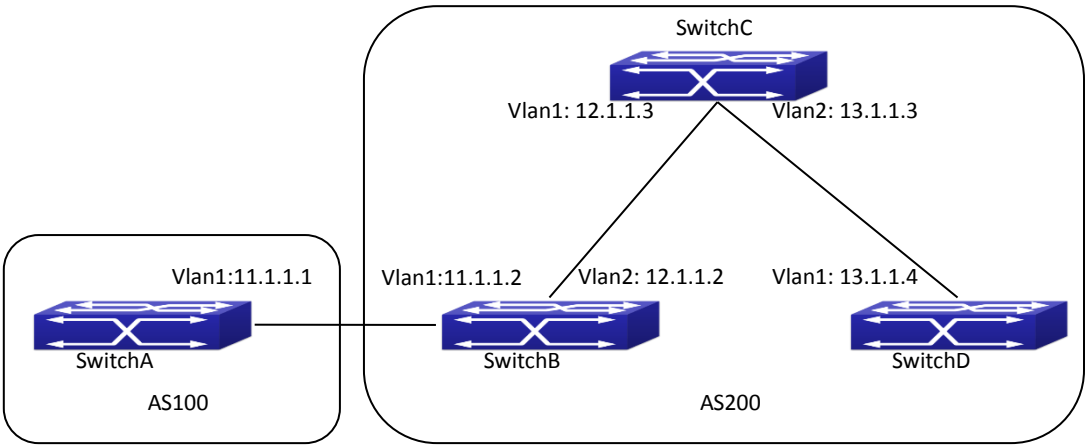


Figure 4-9 BGP Network Topological Map

The configurations of SwitchA are as following:
SwitchA(config)#router bgp 100
SwitchA(config-router-bgp)#neighbor 11.1.1.2 remote-as 200
SwitchA(config-router-bgp)#exit

The configurations of SwitchB are as following:
SwitchB(config)#router bgp 200
SwitchB(config-router-bgp)#network 11.0.0.0
SwitchB(config-router-bgp)#network 12.0.0.0
SwitchB(config-router-bgp)#network 13.0.0.0
SwitchB(config-router-bgp)#neighbor 11.1.1.1 remote-as 100
SwitchB(config-router-bgp)#neighbor 12.1.1.3 remote-as 200
SwitchB(config-router-bgp)#neighbor 13.1.1.4 remote-as 200
SwitchB(config-router-bgp)#exit

The configurations of SwitchC are as following:

SwitchC(config)#router bgp 200
SwitchC(config-router-bgp)#network 12.0.0.0
SwitchC(config-router-bgp)#network 13.0.0.0
SwitchC(config-router-bgp)#neighbor 12.1.1.2 remote-as 200
SwitchC(config-router-bgp)#neighbor 13.1.1.4 remote-as 200
SwitchC(config-router-bgp)#exit

The configurations of SwitchD are as following:
SwitchD(config)#router bgp 200
SwitchD(config-router-bgp)#network 13.0.0.0
SwitchD(config-router-bgp)#neighbor 12.1.1.2 remote-as 200
SwitchD(config-router-bgp)#neighbor 13.1.1.3 remote-as 200
SwitchD(config-router-bgp)#exit

Presently, the connection between SwitchB and SwitchA is EBGP, and other connections with SwitchC and SwitchD are IBGP. SwitchB and SwitchD may have BGP connection without physical connection. But there is a precondition that these two switches must have reachable route to each other. This route can be attained through static route or IGP.

4.5.3.2 Examples 2: configure BGP aggregation

In this sample, configure route aggregation. Firstly, enable command redistribute to redistribute static route to BGP route table:
SwitchB(config)#ip route 193.0.0.0/24 11.1.1.1
SwitchB(config)#router bgp 100
SwitchB(config-router-bgp)#redistribute static
When there is at least one route affiliated to the specified range, the following configuration will create an aggregation route in the BGP route table. The aggregation route will be regarded as the AS from itself. More detailed route information about 193.0.0.0 will be announced.
SwitchB(config#router bgp 100
SwitchB(config-router-bgp)#aggregate 193.0.0.0/16
At the same time, the aggregation command above can be modified as following, then this switch only announce aggregation route 193.0.0.0 and forbid to announce more specified route to all the neighbors.
SwitchB(config-router-bgp)#aggregate 193.0.0.0/16 summary-only

4.5.3.3 Examples 3: configure BGP community attributes

In the following sample, “route map set-community” is used for the outgoing update to neighbor 16.1.1.6. By accessing to route in table 1 to configure special community value to “1111”, other can be announced normally.
Switch(config)#router bgp 100
Switch(config-router-bgp)#neighbor 16.1.1.6 remote-as 200
Switch(config-router-bgp)#neighbor 16.1.1.6 route-map set-community out
Switch(config-router-bgp)#exit

```
Switch(config)#route-map set-community permit 10
Switch(config-route-map)#match address 1
Switch(config-route-map)#set community 1111
Switch(config-route-map)#exit
Switch(config)#route-map set-community permit 20
Switch(config-route-map)#match address 2
Switch(config-route-map)#exit
Switch(config)#access-list 1 permit 11.1.0.0 0.0.255.255
Switch(config)#access-list 2 permit 0.0.0.0 255.255.255.255
Switch(config)#exit
Switch#clear ip bgp 16.1.1.6 soft out
```

In the following sample, configure the MED local preference of the routes from neighbor 16.1.1.6 selectively according to the route community value. All the routes that match the community list will set MED as 2000, community list com1 permits the route with community value “100 200 300”or”900 901” to pass. This route may have other community attributes. All the routes that pass community list com2 will set the local preference as 500. But the route that can’t pass both com1 and com2 will be rejected.

```
Switch(config)#router bgp 100
Switch(config-router-bgp)#neighbor 16.1.1.6 remote-as 200
Switch(config-router-bgp)#neighbor 16.1.1.6 route-map match-community in
Switch(config-router-bgp)#exit
Switch(config)#route-map match-community permit 10
Switch(config-route-map)#match community com1
Switch(config-route-map)#set metric 2000
Switch(config-route-map)#exit
Switch(config)#route-map match-community permit 20
Switch(config-route-map)#match community com2
Switch(config-route-map)#set local-preference 500
Switch(config-route-map)#exit
Switch(config)#ip community-list com1 permit 100 200 300
Switch(config)#ip community-list com1 permit 900 901
Switch(config)#ip community-list com2 permit 88
Switch(config)#ip community-list com2 permit 90
Switch(config)#exit
Switch#clear ip bgp 16.1.1.6 soft out
```

4.5.3.4 Examples 4: configure BGP confederation

The following is the configuration of an AS. As the picture illustrated, SwitchB and SwitchC establish IBGP connection. SwitchD is affiliated to AS 20.SwitchB and SwitchC establish EBGP of inner AS confederation. AS10 and AS20 form AS confederation with the AS number AS200; SwitchA belongs to AS100, SwitchB may create EBGP connection by AS200.

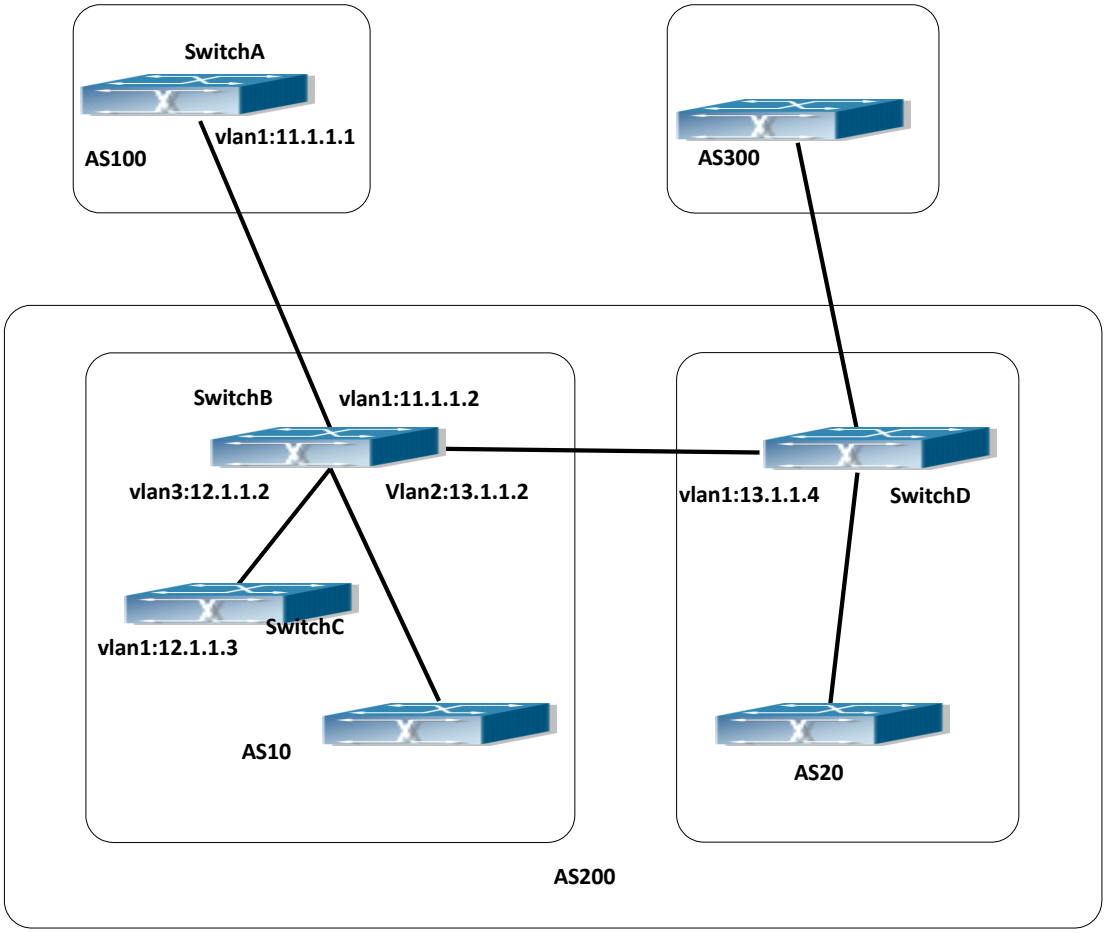


Figure 4-10 Confederation configuring topology

The configurations are as following:

SwitchA:

SwitchA(config)#router bgp 100

SwitchA(config-router-bgp)#neighbor 11.1.1.2 remote-as 200

SwitchB:

SwitchB(config)#router bgp 10

SwitchB(config-router-bgp)#bgp confederation identifier 200

SwitchB(config-router-bgp)#bgp confederation peers 20

SwitchB(config-router-bgp)#neighbor 12.1.1.3 remote-as 10

SwitchB(config-router-bgp)#neighbor 13.1.1.4 remote-as 20

SwitchB(config-router-bgp)#neighbor 11.1.1.1 remote-as 100

SwitchC:

SwitchC(config)#router bgp 10

SwitchC(config-router-bgp)#bgp confederation identifier 200

SwitchC(config-router-bgp)#bgp confederation peers 20

SwitchC(config-router-bgp)#neighbor 12.1.1.2 remote-as 10

SwitchD:

SwitchD(config)#router bgp 20
SwitchD(config-router-bgp)#bgp confederation identifier 200
SwitchD(config-router-bgp)#bgp confederation peers 10
SwitchD(config-router-bgp)#neighbor 13.1.1.2 remote-as 10

4.5.3.5 Examples 5: configure BGP route reflector

The following is the configuration of a route reflector. As the picture illustrated, SwitchA, SwitchB, SwitchC, SwitchD, SWE, SWF and SWG establish IBGP connection which is affiliated to AS100. SwitchC creates EBGP connection with AS200. SwitchA creates EBGP connection with AS300. SwitchC, SwitchD and SWG make route reflectors.

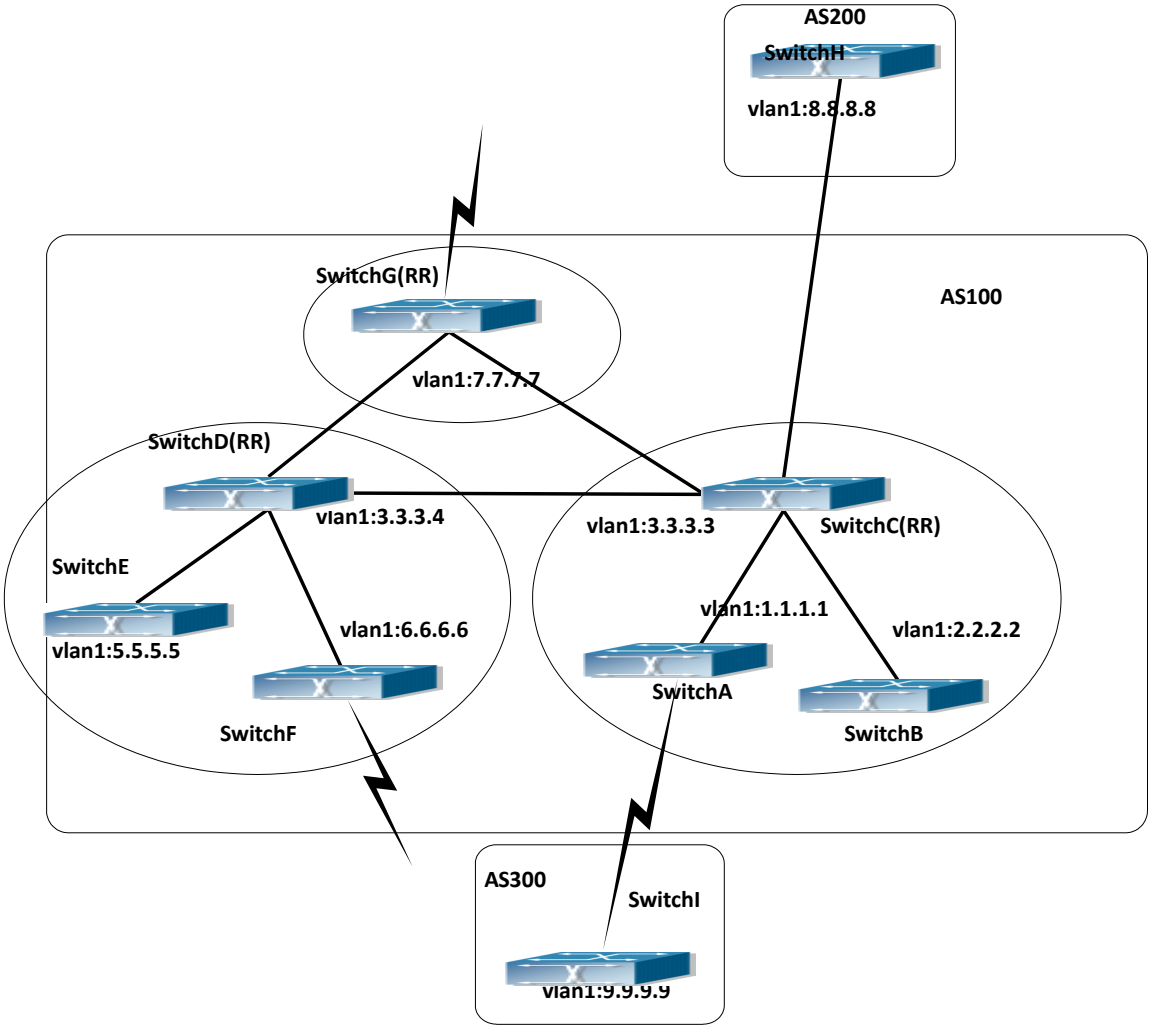


Figure 4-11 the Topological Map of Route Reflector

The configurations are as following:
The configurations of SwitchC:
SwitchC(config)#router bgp 100
SwitchC(config-router-bgp)#neighbor 1.1.1.1 remote-as 100

```
SwitchC(config-router-bgp)#neighbor 1.1.1.1 route-reflector-client
SwitchC(config-router-bgp)#neighbor 2.2.2.2 remote-as 100
SwitchC(config-router-bgp)#neighbor 2.2.2.2 route-reflector-client
SwitchC(config-router-bgp)#neighbor 7.7.7.7 remote-as 100
SwitchC(config-router-bgp)#neighbor 3.3.3.4 remote-as 100
SwitchC(config-router-bgp)#neighbor 8.8.8.8 remote-as 200
```

The configurations of SwitchD:

```
SwitchD(config)#router bgp 100
SwitchD(config-router-bgp)#neighbor 5.5.5.5 remote-as 100
SwitchD(config-router-bgp)#neighbor 5.5.5.5 route-reflector-client
SwitchD(config-router-bgp)#neighbor 6.6.6.6 remote-as 100
SwitchD(config-router-bgp)#neighbor 6.6.6.6 route-reflector-client
SwitchD(config-router-bgp)#neighbor 3.3.3.3 remote-as 100
SwitchD(config-router-bgp)#neighbor 7.7.7.7 remote-as 100
```

The configurations of SwitchA:

```
SwitchA(config)#router bgp 100
SwitchA(config-router-bgp)#neighbor 1.1.1.2 remote-as 100
SwitchA(config-router-bgp)#neighbor 9.9.9.9 remote-as 300
```

The SwitchA at this time needn’t to create IBGP connection with all the switches in the AS100 and could receive BGP route from other switches in the AS.

4.5.3.6 Examples 6: configure MED of BGP

The following is the configuration of a MED. As illustrated, SwitchA is affiliated to AS100, SwitchB is affiliated to AS400, SwitchC and SwitchD belong to AS300.

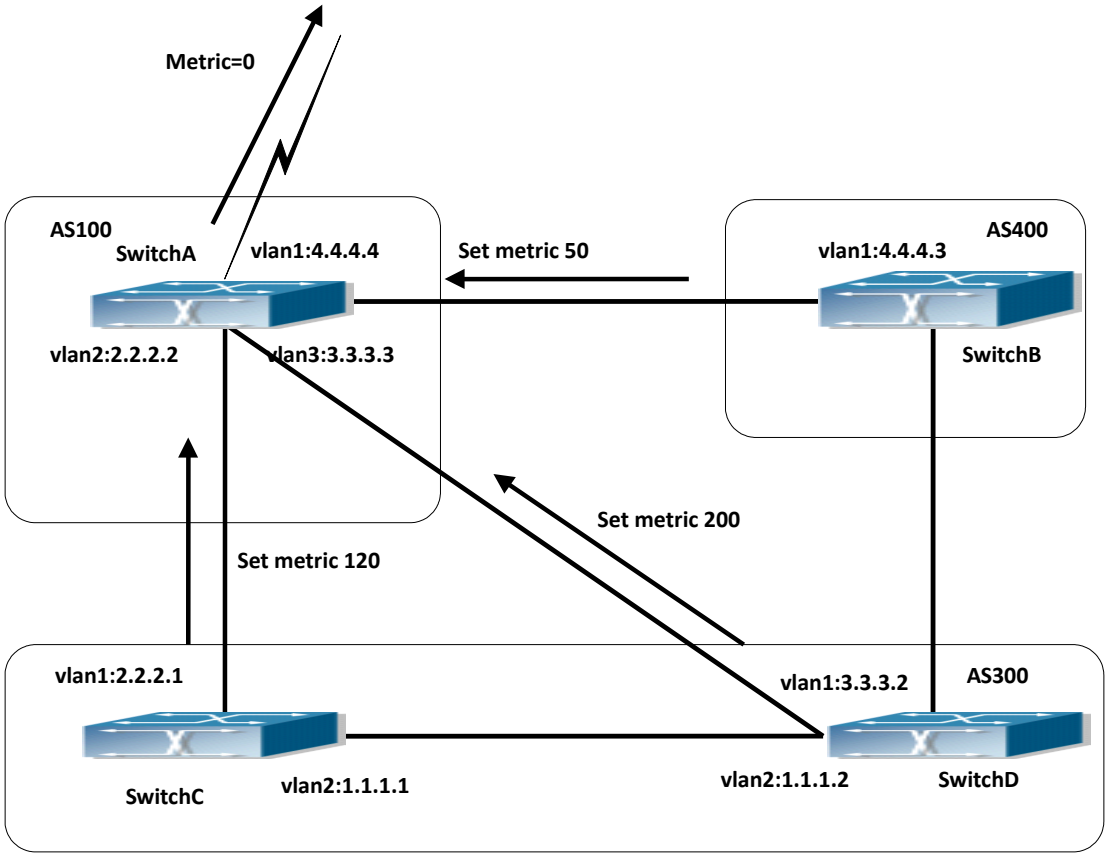


Figure 4-12 MED Configuring Topological Map

The configurations of SwitchA:

SwitchA(config)#router bgp 100

SwitchA(config-router-bgp)#neighbor 2.2.2.1 remote-as 300

SwitchA(config-router-bgp)#neighbor 3.3.3.2 remote-as 300

SwitchA(config-router-bgp)#neighbor 4.4.4.3 remote-as 400

The configurations of SwitchC:

SwitchC(config)#router bgp 300

SwitchC (config-router-bgp)#neighbor 2.2.2.2 remote-as 100

SwitchC (config-router-bgp)#neighbor 2.2.2.2 route-map set-metric out

SwitchC (config-router-bgp)#neighbor 1.1.1.2 remote-as 300

SwitchC (config-router-bgp)#exit

SwitchC (config)#route-map set-metric permit 10

SwitchC (Config-Router-RouteMap)#set metric 120

The configurations of SwitchD

SwitchD (config)#router bgp 300

SwitchD (config-router-bgp)#neighbor 3.3.3.3 remote-as 100

SwitchD (config-router-bgp)#neighbor 3.3.3.3 route-map set-metric out

SwitchD (config-router-bgp)#neighbor 1.1.1.1 remote-as 300

SwitchD (config-router-bgp)#exit

SwitchD (config)#route-map set-metric permit 10
SwitchD (Config-Router-RouteMap)#set metric 200

The configurations of SwitchB
SwitchB (config)#router bgp 400
SwitchB (config-router-bgp)#neighbor 4.4.4.4 remote-as 100
SwitchB (config-router-bgp)#neighbor 4.4.4.4 route-map set-metric out
SwitchB (config-router-bgp)#exit
SwitchB (config)#route-map set-metric permit 10
SwitchB (Config-Router-RouteMap)#set metric 50

After the configuration above, SwitchB, SwitchC and SwitchD are assumed to send a route 12.0.0.0 to SwitchA. According to the comparison of BGP route strategy; there is an assumption that the routes sent by the three switches above have the same attribute value before the comparison of metric attribute. At this time, the route with lower value is the better route. But the comparison of metric attribute will only be done with the routes from the same AS. For SwitchA, the routes passed SwitchC are preferable to the one passed SwitchD. Because SwitchC and SwitchB are not located in the same AS, the SwitchA will not do metric comparison between the two switches. If the metric comparison between different AS is needed, the command" bgp always-compare-med" will be used. If this command is configured, the routes passed SwitchB are the best to SwitchA. At this time, the following command may be added on SwitchA:
SwitchA (config-router-bgp)#bgp always-compare-med

4.5.3.7 Examples 7: example of BGP VPN

For the configuration of MPLS VPN, BGP is part of the core routing system and it is also an important utility to support ILM and FTN entries on the edge devices. For switch, the BGP protocol together with the LDP protocol, constructs the foundation of the MPLS VPN application. The LDP protocol works at the WLAN side and for the routers which are not on the edge of the network, the BGP protocol does not function.

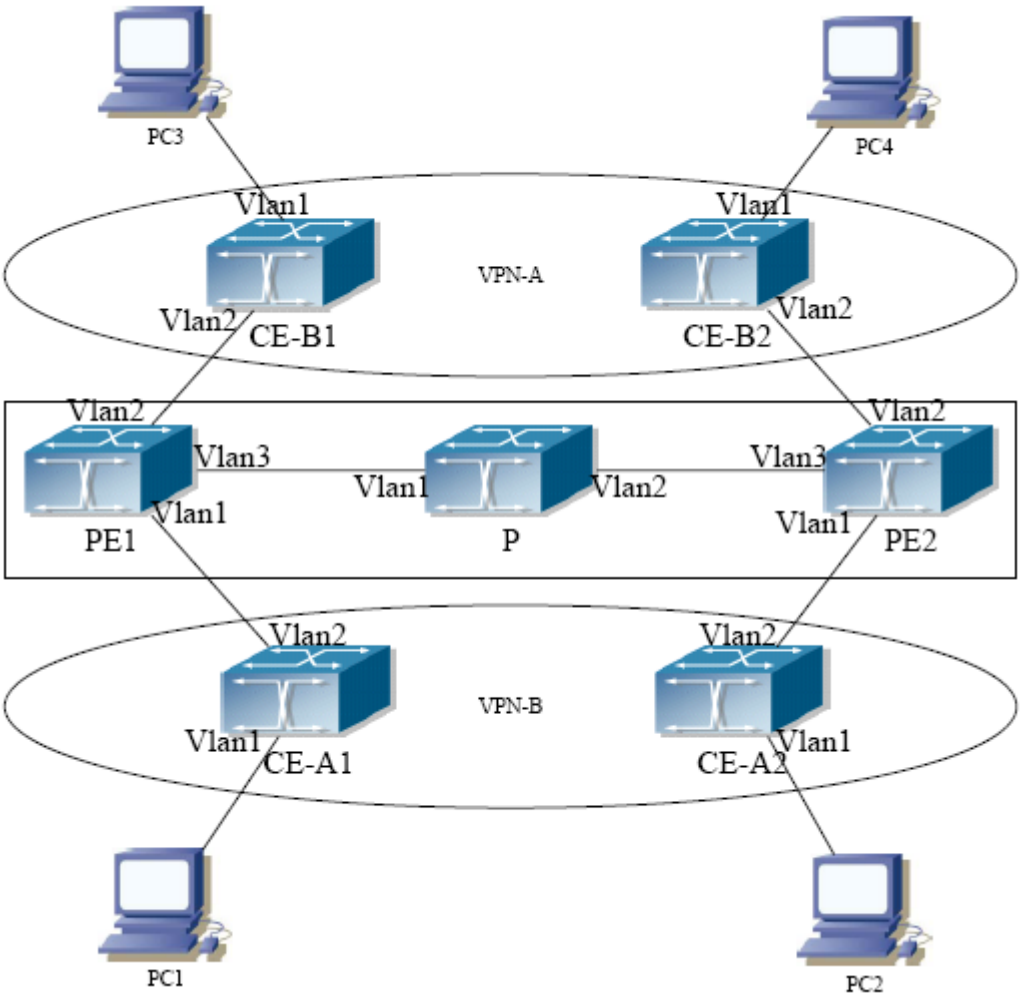


Figure 4-13 Example of MPLS VPN

As the figure shows, for a typical MPLS VPN application, the public network region consists of PE1, P and PE2, which MPLS is applied for packet transmission. VPN-A consists of CE-A1 and CE-A2, and VPN-B consists of CE-B1 and CE-B2. These two VPNs are isolated from each other. PE1 and PE2 are edge routers which are provided by the operators. CE-A1, CE-A2, CE-B1 and CE-B2 are the access switches on the user side. PC1-PC4 indicate the network users. BGP runs at both the public and private network region. For the public network region, VPN routing should be supported and the LOOPBACK interface should be used for connections.

The sample configurations are listed as below.

Configurations on CE-A1：

```
CE-A1#config
CE-A1(config)#interface vlan 2
CE-A1(config-if-Vlan2)#ip address 192.168.101.2 255.255.255.0
CE-A1(config-if-Vlan2)#exit
CE-A1(config)#interface vlan 1
CE-A1(config-if-Vlan2)#ip address 10.1.1.1 255.255.255.0
CE-A1(config-if-Vlan2)#exit
CE-A1(config)#router bgp 60101
```

```
CE-A1(config-router)#neighbor 192.168.101.1 remote-as 100
CE-A1(config-router)#exit
```

Configurations on CE-A2：.

```
CE-A2#config
CE-A2(config)#interface vlan 2
CE-A2(config-if-Vlan2)#ip address 192.168.102.2 255.255.255.0
CE-A2(config-if-Vlan2)#exit
CE-A2(config)#interface vlan 1
CE-A2(config-if-Vlan2)#ip address 10.1.2.1 255.255.255.0
CE-A2(config-if-Vlan2)#exit
CE-A2(config)#router bgp 60102
CE-A2(config-router)#neighbor 192.168.102.1 remote-as 100
CE-A2(config-router)#exit
```

Configurations on CE-B1：.

```
CE-B1#config
CE-B1(config)#interface vlan 2
CE-B1(config-if-Vlan2)#ip address 192.168.201.2 255.255.255.0
CE-B1(config-if-Vlan2)#exit
CE-B1(config)#interface vlan 1
CE-B1(config-if-Vlan2)#ip address 20.1.1.1 255.255.255.0
CE-B1(config-if-Vlan2)#exit
CE-B1(config)#router bgp 60201
CE-B1(config-router)#neighbor 192.168.201.1 remote-as 100
CE-B1(config-router)#exit
```

Configurations on CE-BE2：.

```
CE-B2#config
CE-B2(config)#interface vlan 2
CE-B2(config-if-Vlan2)#ip address 192.168.202.2 255.255.255.0
CE-B2(config-if-Vlan2)#exit
CE-B2(config)#interface vlan 1
CE-B2(config-if-Vlan2)#ip address 20.1.2.1 255.255.255.0
CE-B2(config-if-Vlan2)#exit
CE-B2(config)#router bgp 60202
CE-B2(config-router)#neighbor 192.168.202.1 remote-as 100
CE-B2(config-router)#exit
```

Configurations on PE1：

```
PE1#config
PE1(config)#ip vrf VRF-A
PE1(config-vrf)#rd 100:10
PE1(config-vrf)#route-target both 100:10
```

```
PE1(config-vrf)#exit
PE1(config)#ip vrf VRF-B
PE1(config-vrf)#rd 100:20
PE1(config-vrf)#route-target both 100:20
PE1(config-vrf)#exit
PE1(config)#interface vlan 1
PE1(config-if-Vlan1)#ip vrf forwarding VRF-A
PE1(config-if-Vlan1)#ip address 192.168.101.1 255.255.255.0
PE1(config-if-Vlan1)#exit
PE1(config)#interface vlan 2
PE1(config-if-Vlan2)#ip vrf forwarding VRF-B
PE1(config-if-Vlan2)#ip address 192.168.201.1 255.255.255.0
PE1(config-if-Vlan2)#exit
PE1(config)#interface vlan 3
PE1(config-if-Vlan3)#ip address 202.200.1.2 255.255.255.0
PE1(config-if-Vlan3)#label-switching
PE1(config-if-Vlan3)#exit
PE1(config)#interface loopback 1
PE1(Config-if-Loopback1)# ip address 200.200.1.1 255.255.255.255
PE1(config-if-Vlan3)#exit
PE1(config)#router bgp 100
PE1(config-router)#neighbor 200.200.1.2 remote-as 100
PE1(config-router)#neighbor 200.200.1.2 update-source 200.200.1.1
PE1(config-router)#address-family vpnv4 unicast
PE1(config-router-af)#neighbor 200.200.1.2 activate
PE1(config-router-af)#exit-address-family
PE1(config-router)#address-family ipv4 vrf VRF-A
PE1(config-router-af)# neighbor 192.168.101.2 remote-as 60101
PE1(config-router-af)#exit-address-family
PE1(config-router)#address-family ipv4 vrf VRF-B
PE1(config-router-af)# neighbor 192.168.201.2 remote-as 60201
PE1(config-router-af)#exit-address-family
```

Configurations on PE2：

```
PE2#config
PE2(config)#ip vrf VRF-A
PE2(config-vrf)#rd 100:10
PE2(config-vrf)#route-target both 100:10
PE2(config-vrf)#exit
PE2(config)#ip vrf VRF-B
PE2(config-vrf)#rd 100:20
PE2(config-vrf)#route-target both 100:20
PE2(config-vrf)#exit
PE2(config)#interface vlan 1
```

```
PE2(config-if-Vlan1)#ip vrf forwarding VRF-A
PE2(config-if-Vlan1)#ip address 192.168.102.1 255.255.255.0
PE2(config-if-Vlan1)#exit
PE2(config)#interface vlan 2
PE2(config-if-Vlan2)#ip vrf forwarding VRF-B
PE2(config-if-Vlan2)#ip address 192.168.202.1 255.255.255.0
PE2(config-if-Vlan2)#exit
PE2(config)#interface vlan 3
PE2(config-if-Vlan3)#ip address 202.200.2.2 255.255.255.0
PE2(config-if-Vlan3)#label-switching
PE2(config-if-Vlan3)#exit
PE2(config)#interface loopback 1
PE2(Config-if-Loopback1)# ip address 200.200.1.2 255.255.255.255
PE2(config-if-Vlan3)#exit
PE2(config)#router bgp 100
PE2(config-router)#neighbor 200.200.1.1 remote-as 100
PE2(config-router)#address-family vpnv4 unicast
PE2(config-router-af)#neighbor 200.200.1.1 activate
PE2(config-router-af)#exit-address-family
PE2(config-router)#address-family ipv4 vrf VRF-A
PE2(config-router-af)# neighbor 192.168.102.2 remote-as 60102
PE2(config-router-af)#exit-address-family
PE2(config-router)#address-family ipv4 vrf VRF-B
PE2(config-router-af)# neighbor 192.168.202.2 remote-as 60202
PE2(config-router-af)#exit-address-family
```

The sample configurations which are listed above is the most typical one. To enable communication between VRF, the route-target should be modified. And if the BGP AS number duplicates for the ends, the "**neighbor *<ip-addr>* as-override"** command should be configured to avoid the duplication of AS numbers.

Also, only BGP related configuration are listed above, to run LDP on the public network region, please refer to the LDP configuration sample.

4.5.4 BGP Troubleshooting

In the process of configuring and implementing BGP protocol, physical connection, configuration false probably leads to BGP protocol doesn't work. Therefore, the customers should give their attention to points as follow:

- ☞ First of all, to ensure correct physical connection;
- ☞ Secondly, to ensure interface and link protocol are UP (execute **show interface** instruction);
- ☞ And startup BGP protocol (use **router bgp** command), configure affiliated IBGP and EBGP neighbors (use **neighbor remote-as** command).

- ☞ Notice BGP protocol itself can't detect route, needs to import other routes to create BGP route. Only it enables these routes to announce IBGP and EBGP neighbors by importing routes. Direct-link routes, static route, and IGP route (RIP and OSPF) are included in these imported routes. **network** and **redistribute (BGP)** command are the ways of imported routes.
- ☞ For BGP, pay attention to the difference between the behaviors of IBGP and EBGP.
- ☞ After configuration finishes, the command of **show ip bgp summary** can be used to observe neighbor's connections, so that all of the neighbors keep BGP connection situation. And use **show ip bgp** command to observe BGP routing table.
- ☞ If BGP routing problem still can't be solved by debugging, please use debug instructions like **debug ip bgp** packet/events etc, and copy DEBUG information in 3 minutes, then send them to ourTechnology Service Center.

4.6 IPv4 Black Hole Routing

4.6.1 Introduction to Black Hole Routing

Black Hole Routing is a special kind of static routing which drops all the datagrams that match the routing rule.

4.6.2 IPv4 Black Hole Routing Configuration Task

1. Configure IPv4 Black Hole Routing

1. Configure IPv4 Black Hole Routing

Command	Explaination
Global Configuration Mode	
ip route {*<ip-prefix> <mask>*|*<ip-prefix>/<prefix-length>*} null0 [*<distance>*] **no ip route {*<ip-prefix> <mask>*|*<ip-prefix>/<prefix-length>*} null0**	To configure the static Black Hole Routing. The no form of this command will remove the specified Black Hole Routing configuration.

4.6.3 Black Hole Routing Configuration Exmaples

Example 1: IPv4 Black Hole Routing function.

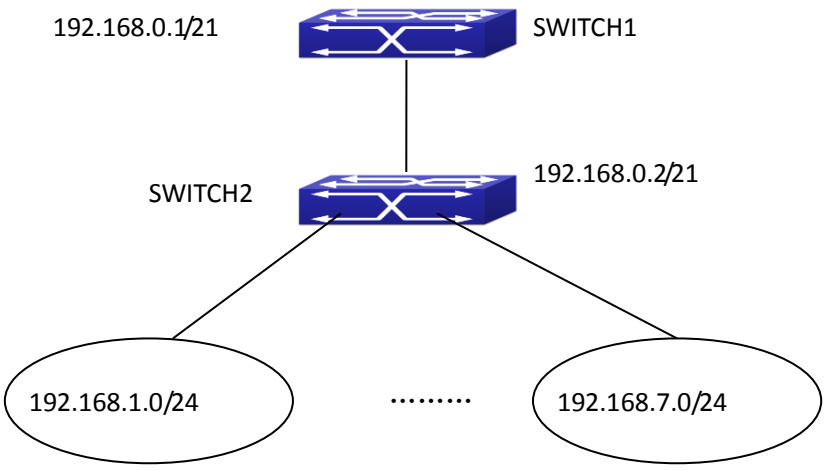


Figure 4-14 IPv4 Black Hole Routing Configuration Example

As it is shown in the figure, in Switch 2, eight in all interfaces are configured as Layer 3 VLAN interfaces for access interfaces. The network addresses are 192.168.1.0/24 ~ 192.268.7.0/24. A default routing is configured on Switch 2 to connect to Switch 1. And a backward default routing is configured on Switch 1 to Switch 2, whose network address is 192.168.0.0/21. Commonly, this configuration will work well. However, if one of the Layer 3 interfaces in Switch 2 goes down, for example, the interface belonged to 192.168.1.0/24. When datagrams arrives at VLAN1 in Switch 2, there will be no routing rules for these datagrams. The switch then will forward these datagrams according to the default routing, back to Switch 1. When Switch 1 receives these datagrams, it will forward them back to Switch 2. Thus, loopback exists. To solve this problem, Black Hole Routing can be introduced on Switch 2.

ip route 192.168.0.0/21 null0 50

Then Switch 2 will drop the datagrams from interface VLAN1 that match the Black Hole Routing rule. And loopback routing is prevented.

Configuration steps are listed as below:

Switch#config

Switch(config)#ip route 192.168.0.0/21 null0 50

4.6.4 Black Hole Routing Troubleshooting

When configuring the Black Hole Routing function, the configuration may not work due to some reasons such as incorrect network address mask, and incorrect management distance. Attention should be paid to the following items:

- IPv6 should be enabled before IPv6 Black Hole Routing can work.
- It is suggested that the length of the network address mask should be longer than that of normal routing configuration, in order to prevent the Black Hole Routing from intervening other routing configuration.
- When the network address mask of Black Hole Routing configuration is the same with some

other configuration, it is suggested that the distance of Black Hole Routing is set lower.

For problems that cannot be fixed through above methods, please issue the command show ip route distance and show ip route fib, and show l3. And copy and paste the output of the commands, and send to the technical service center of our company.

4.7 GRE

4.7.1 Introduction to GRE Tunnel

GRE (General Routing-protocol Encapsulation) was referred to IETF by Cisco and Net-smiths companies in 1994, in RFC1701 and RFC1702. At present, the network devices of the most manufacturers support the GRE tunnel protocol. GRE set how to encapsulate a kind of network protocol in other kind of network protocol. The GRE tunnel is defined by the source IP address and the destination IP address of two ends, it allows that users can use IP packets to encapsulate IP, IPX, AppleTalk packets, and supports all routing protocols (such as RIP2, OSPF, etc). By GRE, users can use the public IP network to connect IPX network, AppleTalk network, and use the reserved address to connect through the internet work, or hide the IP address of the corporation network to the public network. GRE only provides the encapsulation of the data packets, and has no encryption function to avoid the network interception and attack. Therefore, it is used with IPsec usually in the actual environment, and IPsec provides the data encryption to supply very good security for users.

The most use of GRE protocol: the internal protocol encapsulation and the private address encapsulation of the enterprise. In our nation, the enterprise network almost adopts TCP/IP protocol, so there are fewer requirements to the internal protocol encapsulation. The enterprise use GRE almost for encapsulating the private address .In our switches, the GRE is used mainly for the transition of the network protocol (include IPv6 OVER IPv4 and IPv4 OVER IPv6).

RFC1701, 1702, 2784 are referred for the implement.

4.7.2 GRE Tunnel Basic Configuration

GRE Tunnel Configuration Task List:

1. Configure tunnel mode
    1) Configure tunnel mode as GREv4 tunnel
2. Configure the source address and the destination address of GRE tunnel
    1) Configure the source address as the IPv4 address for GRE tunnel
    2) Configure the destination address as the IPv4 address for GRE tunnel
3. Configure the interface address of GRE tunnel
    1) Configure the IPv4 address of GRE tunnel interface
    2) Configure the IPv6 address of GRE tunnel interface
4. Configure the egress interface of the static route as GRE tunnel

1) Configure the egress interface of the IPv4 static route to GRE tunnel
2) Configure the egress interface of the IPv6 static route to GRE tunnel

1. Configure tunnel mode

Command	Explanation
Tunnel interface configuration mode	
tunnel mode gre ip **no tunnel mode**	Configure the tunnel mode as GREv4 tunnel. After the data packet is encapsulated with GRE, it has a head of IPv4 packets, and passes the IPv4 network.

2. Configure the source address and the destination address of GRE tunnel

Command	Explanation
Tunnel interface configuration mode	
tunnel source *<ipv4-address>* **no tunnel source**	Configure the source address as the IPv4 address for GRE tunnel.
tunnel destination *<ipv4-address>* **no tunnel destination**	Configure the destination address as the IPv4 address for GRE tunnel.

Command	Explanation
Tunnel interface configuration mode	
tunnel source {*<ipv6-address>* | *<ipv4-address>*} **no tunnel source**	Configure the source address as the IPv6 or IPv4 address for GRE tunnel.
tunnel destination {*<ipv6-address>* | *<ipv4-address>*} **no tunnel destination**	Configure the destination address as the IPv6 or IPv4 address for GRE tunnel.

3. Configure the interface address of GRE tunnel

Command	Explanation
Tunnel interface configuration mode	
ip address *<ipv4-address> <mask>* **no ip address *<ipv4-address> <mask>***	Configure the IPv4 address of GRE tunnel interface.
ipv6 address *<ipv6-address/prefix>* **no ipv6 address *<ipv6-address/prefix>***	Configure the IPv6 address of GRE tunnel interface.

4. Configure the egress interface of the static route as GRE tunnel

Command	Explanation

Global Mode	
ip route *<ipv4-address/mask>* tunnel *<ID>* **no ip route *<ipv4-address/mask>* tunnel *<ID>***	Configure the egress interface of the IPv4 static route to GRE tunnel.
ipv6 route *<ipv6-address/prefix>* tunnel *<ID>* **no ipv6 route *<ipv6-address/prefix>* tunnel *<ID>***	Configure the egress interface of the IPv6 static route to GRE tunnel.

4.7.3 Example of GRE Tunnel

The Typical Example of GRE Tunnel:

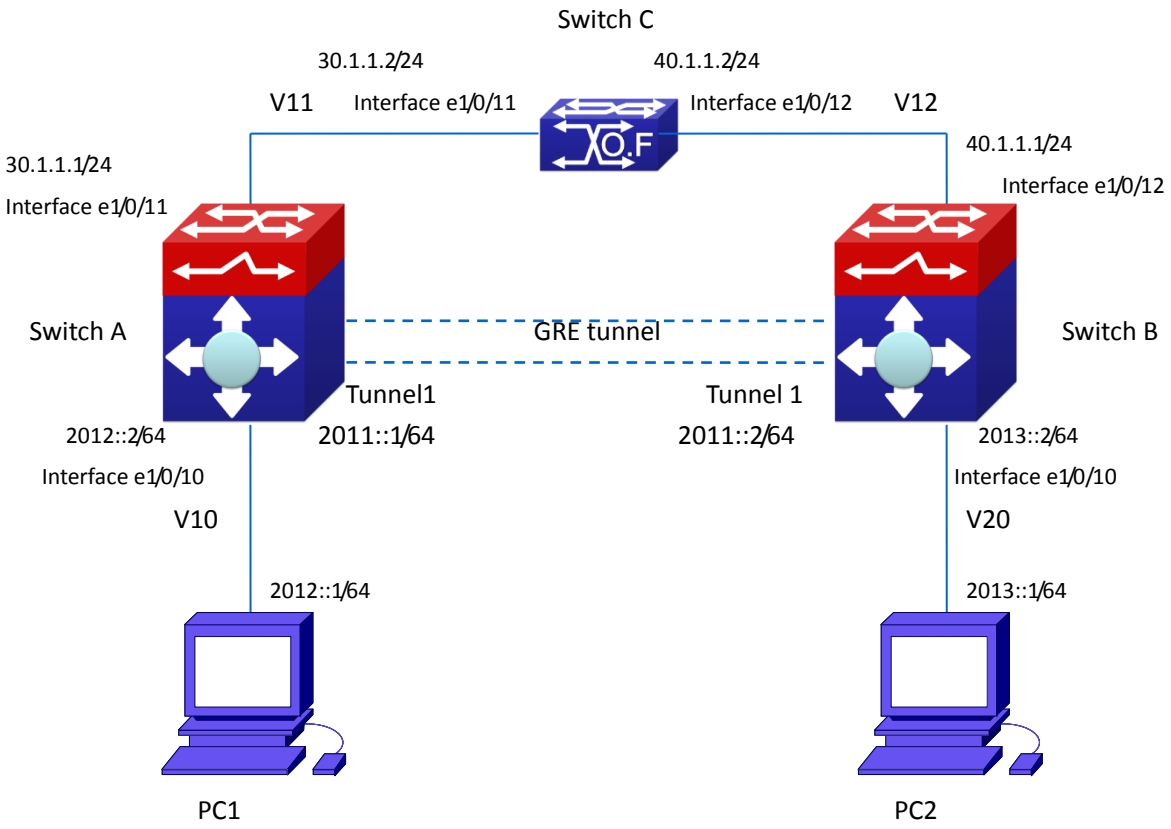


Figure 4-15 the typical network of IPv6 over IPv4 GRE tunnel character

Configuration summary:

- Configure IPv4 network, and ensure IPv4 connectivity.
- Configure the tunnel interface, and the interface that connect PC.
- Configure the tunnel parameters and enable the tunnel interface.
- Enable OSPF route protocol to forward the data between PC1 and PC2 through the tunnel.

Configuration steps

Instruction: the topology environment of this chapter may be different to the actual environment. To ensure the effect of the configuration, please make sure the current configuration of the device does not conflict with the following configuration.

(1) The configuration of device A

1. The configuration step

☞ Create the interface VLAN 11 and its address.

```
SwitchA(config)#vlan 11
SwitchA(config-vlan11)#switchport interface ethernet 1/0/11
SwitchA(config-vlan11)#exit
SwitchA(config)#interface vlan 11
SwitchA(config-if-vlan11)#ip address 30.1.1.1/24
```

☞ Configure the IPv4 static route to switch B from interface Vlan11.

```
SwitchA(config)#ip route 40.1.1.1/24 30.1.1.2
```

☞ Configure tunnel interface: source, destination and type. After the tunnel is enabled, its source address and destination address can not be modified except the source address as the layer 3 interface.

```
SwitchA(config)#interface tunnel 1
SwitchA(config-if-tunnel1)# tunnel source 30.1.1.1
SwitchA(config-if-tunnel1)# tunnel destination 40.1.1.1
SwitchA(config-if-tunnel1)# tunnel mode gre ip
SwitchA#show gre tunnel
name        mode        source        destination
Tunnel1     gre ip      30.1.1.1      40.1.1.1
```

The configuration of GRE tunnel is successful.

☞ Configure the IPv6 address of the tunnel interface. Only one interface address can be configured for the tunnel interface, and this limitation will also be used to other tunnels, such as configure tunnel, 6to4, isatap.

Notice: the tunnel must stays in active state when configuring IPv4 address, but configure IPv6 address is different to IPv4 address.

```
SwitchA (config-if-tunnel1)#ipv6 address 2011::1/64
```

☞ Configure the interface VLAN10 and its address.

```
SwitchA(config)#vlan 10
SwitchA(config-vlan10)#switchport interface ethernet 1/0/10
SwitchA(config-vlan10)#exit
SwitchA(config)#interface vlan 10
SwitchA(config-if-vlan10)# ipv6 address 2012::2/64
SwitchA(config-if-vlan10)#exit
```

☞ Configure OSPF routing protocol.

```
SwitchA(config)#router ospf
SwitchA(config-router)#router-id 1.1.1.1
SwitchA(config-router)#network 30.1.1.1/24 area 0
SwitchA(config-router)#exit
```

(2) The configuration of device B

1. The configuration step

☞ Create the interface VLAN 12 and its address.

```
SwitchA(config)#vlan 12
SwitchA(config-vlan12)#switchport interface ethernet 1/0/12
SwitchA(config-vlan12)#exit
SwitchA(config)#interface vlan 12
SwitchA(config-if-vlan12)#ip address 40.1.1.1/24
SwitchA(config-if-vlan12)#exit
SwitchA(config)#
```

☞ Configure the IPv4 static route to Switch A from interface Vlan12.

```
SwitchA(config)#ip route 30.1.1.1/24 40.1.1.2
```

☞ Configure tunnel interface: source, destination and type.

```
SwitchA(config)#interface tunnel 1
SwitchA(config-if-tunnel1)# tunnel source 40.1.1.1
SwitchA(config-if-tunnel1)# tunnel destination 30.1.1.1
SwitchA(config-if-tunnel1)# tunnel mode gre ip
SwitchA#show gre tunnel
name        mode      source        destination
Tunnel1     gre ip    40.1.1.1      30.1.1.1
```

The configuration of GRE tunnel is successful.

☞ Configure the IPv6 address of the tunnel interface. To run OSPF routing protocol, the interface address must be configured.

```
SwitchA (config-if-tunnel1)#ipv6 address 2011::2/64
```

☞ Configure the interface VLAN20 and its address.

```
SwitchA(config)#vlan 20
SwitchA(config-vlan20)#switchport interface ethernet 1/0/10
SwitchA(config-vlan20)#exit
SwitchA(config)#interface vlan 20
SwitchA(config-if-vlan20)# ipv6 address 2013::2/64
SwitchA(config-if-vlan20)#exit
SwitchA(config)#
```

☞ Configure OSPF routing protocol.

```
SwitchA(config)#router ospf
SwitchA(config-router)#router-id 1.1.1.2
SwitchA(config-router)#network 40.1.1.0/24 area 0
SwitchA(config-router)#exit
SwitchA(config)#
```

(3) The configuration of device C

1. The configuration step

☞ Create the interface VLAN 11 and its address.

SwitchA(config)#vlan 11
SwitchA(config-vlan11)#switchport interface ethernet 1/0/11
SwitchA(config-vlan11)#exit
SwitchA(config)#interface vlan 11
SwitchA(config-if-vlan11)#ip address 30.1.1.2/24

☞ Create the interface VLAN 12 and its address
SwitchA(config)#vlan 12
SwitchA(config-vlan12)#switchport interface ethernet 1/0/12
SwitchA(config-vlan12)#exit
SwitchA(config)#interface vlan 12
SwitchA(config-if-vlan12)#ip address 40.1.1.2/24
SwitchA(config-if-vlan12)#exit

(4) The configuration of PC

☞ Configure the IP address of PC1 and the default gateway.
PC1: the IP address: 2012::1/64, the default gateway: 2012::2
PC2: the IP address: 2013::1/64, the default gateway: 2013::2

4.7.4 Example of GRE Tunnel Quotes Loopback Group

Introduction to loopback group

When the cards with different types are inserted to a device, implement the redirection between the different cards through loopback function, such as: the cards whether support GRE tunnel are inserted, so the GRE tunnel data received by the cards (do not support GRE tunnel) to be processed to the supporting GRE tunnel cards. A loopback group uses the (one or several) card's ports without any configuration. To increase the redirection bandwidth between the cards, join many ports in a loopback group to implement the load balance.

Example of GRE tunnel quotes loopback group

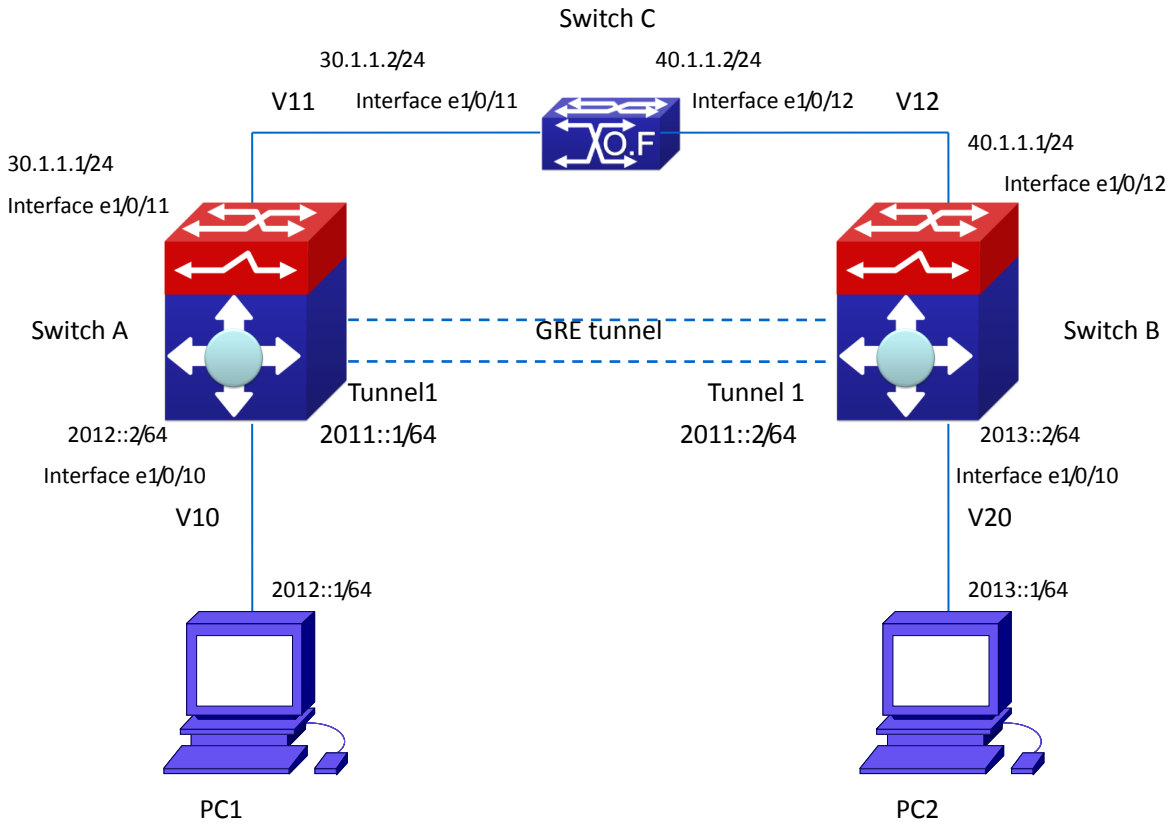


Figure 4-16 GRE tunnel quotes loopback group topology

Introduction to loopback group topology

IPv4 network between SwitchA and SwitchB, PC1 and PC2 at IPv6 network, so PC1 must pass IPv4 network between SwitchA and SwitchB to communicate with PC2 through GRE tunnel. The cards with different types may be inserted to SwitchA, card 1 does not support GRE tunnel, card 3 supports GRE tunnel, so the data received by card 1 needs to be processed to card 3 through loopback group function.

Configuration summary:

- Configure IPv4 network, and ensure IPv4 connectivity.
- Configure the tunnel interface, and the interface that connect PC.
- Configure the tunnel parameters and enable the tunnel interface.
- Configure loopback group, join port 1/0/12 of the card 3 in this loopback group and enable the tunnel to quote this loopback group.
- Enable OSPF route protocol to forward the data between PC1 and PC2 through the tunnel.

Configuration steps

Instruction: the topology environment of this chapter may be different to the actual environment. To ensure the effect of the configuration, please make sure the current configuration of the device does not conflict with the following configuration.

(1) The configuration of device A

1. The configuration step

☞ Create the interface VLAN 11 and its address.

SwitchA(config)#vlan 11
SwitchA(config-vlan11)#switchport interface ethernet 1/0/11
SwitchA(config-vlan11)#exit
SwitchA(config)#interface vlan 11
SwitchA(config-if-vlan11)#ip address 30.1.1.1/24

☞ Configure the IPv4 static route to switch B from interface Vlan11.

SwitchA(config)#ip route 40.1.1.1/24 30.1.1.2

☞ Configure tunnel interface: source, destination and type.

SwitchA(config)#interface tunnel 1
SwitchA(config-if-tunnel1)# tunnel source 30.1.1.1
SwitchA(config-if-tunnel1)# tunnel destination 40.1.1.1
SwitchA(config-if-tunnel1)# tunnel mode gre ip
SwitchA#show gre tunnel

name	mode	source	destination
Tunnel1	gre ipv6	30.1.1.1	40.1.1.1

The configuration of GRE tunnel is successful.

☞ Configure the IPv6 address of the tunnel interface. To run OSPF routing protocol, the interface address must be configured.

SwitchA (config-if-tunnel1)#ipv6 address 2011::1/64

☞ Configure the interface VLAN10 and its address.

SwitchA(config)#vlan 10
SwitchA(config-vlan10)#switchport interface ethernet 1/0/10
SwitchA(config-vlan10)#exit
SwitchA(config)#interface vlan 10
SwitchA(config-if-vlan10)# ipv6 address 2012::2/64
SwitchA(config-if-vlan10)#exit

☞ Configure loopback group and enable the tunnel to quote this loopback group.

SwitchA (config)#loopback-group 1
SwitchA (config-if-ethernet1/0/12)#loopback-group 1
SwitchA (config-if-tunnel1)# loopback-group 1

☞ Configure OSPF routing protocol.

SwitchA(config)#router ospf
SwitchA(config-router)#router-id 1.1.1.1
SwitchA(config-router)#network 30.1.1.0/24 area 0
SwitchA(config-router)#exit

(2) The configuration of device B

1. The configuration step

☞ Create the interface VLAN 12 and its address.

SwitchA(config)#vlan 12
SwitchA(config-vlan12)#switchport interface ethernet 1/0/12
SwitchA(config-vlan12)#exit

```
SwitchA(config)#interface vlan 12
SwitchA(config-if-vlan11)#ip address 30.1.1.2/24
SwitchA(config-if-vlan12)#exit
SwitchA(config)#
```

☞ Configure the IPv4 static route to Switch A from interface Vlan12.

```
SwitchA(config)#ip route 30.1.1.1/24 40.1.1.2
```

☞ Configure tunnel interface: source, destination and type.

```
SwitchA(config)#interface tunnel 1
SwitchA(config-if-tunnel1)# tunnel source 40.1.1.1
SwitchA(config-if-tunnel1)# tunnel destination 30.1.1.1
SwitchA(config-if-tunnel1)# tunnel mode gre ip
SwitchA#show gre tunnel
name        mode        source        destination
Tunnel1     gre ipv6    40.1.1.1      30.1.1.1
```

The configuration of GRE tunnel is successful.

☞ Configure the IPv6 address of the tunnel interface. To run OSPF routing protocol, the interface address must be configured.

```
SwitchA (config-if-tunnel1)#ipv6 address 2011::2/64
```

☞ Configure the interface VLAN20 and its address.

```
SwitchA(config)#vlan 20
SwitchA(config-vlan20)#switchport interface ethernet 1/0/10
SwitchA(config-vlan20)#exit
SwitchA(config)#interface vlan 20
SwitchA(config-if-vlan20)# ipv6 address 2013::2/64
SwitchA(config-if-vlan20)#exit
SwitchA(config)#
```

☞ Configure OSPF routing protocol.

```
SwitchA(config)#router ospf
SwitchA(config-router)#router-id 1.1.1.2
SwitchA(config-router)#network 40.1.1.0/24 area 0
SwitchA(config-router)#exit
SwitchA(config)#
```

(3) The configuration of device C

1. The configuration step

☞ Create the interface VLAN 11 and its address.

```
SwitchA(config)#vlan 11
SwitchA(config-vlan11)#switchport interface ethernet 1/0/11
SwitchA(config-vlan11)#exit
SwitchA(config)#interface vlan 11
SwitchA(config-if-vlan11)#ip address 30.1.1.2/24
```

☞ Create the interface VLAN 12 and its address

```
SwitchA(config)#vlan 12
```

```
SwitchA(config-vlan12)#switchport interface ethernet 1/0/12
SwitchA(config-vlan12)#exit
SwitchA(config)#interface vlan 12
SwitchA(config-if-vlan12)#ip address 40.1.1.2/24
SwitchA(config-if-vlan12)#exit
```

(4) The configuration of PC

- Configure the IP address of PC1 and the default gateway.
  PC1: the IP address: 2012::1/64, the default gateway: 2012::2
  PC2: the IP address: 2013::1/64, the default gateway: 2013::2

4.7.5 GRE Tunnel Troubleshooting

If there is any problem happens when using GRE tunnel, please check whether the problem is caused by the following reasons:

- Check the configuration, whether the source and the destination addresses of the tunnel are correctly configured, whether the tunnel mode (tunnel mode gre {ip | ipv6}) is configured correctly.
- Check the static route that with the GRE tunnel interface as the next hop interface.
- Whether the connection is normal between the switches, use debug gre {packet | event | all} to check whether the switch receives and processes the GRE packets correctly.

4.8 ECMP

4.8.1 Introduction to ECMP

ECMP (Equal-cost Multi-path Routing) works in the network environment where there are many different links to arrive at the same destination address. If using the traditional routing technique, only a link can be used to send the data packets to the destination address, other links at the backup state or the invalidation state, and it needs some times to process the mutual switchover under the static routing environment. However, ECMP protocol can use multi-links under such network environment, it not only implements the load balance, increases the transport bandwidth, but also can completely backup the data transport of the invalidation links without delay and packet loss.

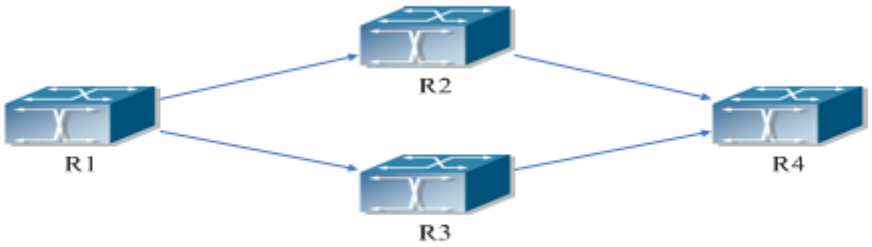


Figure 4-17 the application environment of ECMP

As it is shown in the figure, there are two paths can be selected from R1 to R4, they are R1-R2-R4 and R1-R3-R4. If the route type and the cost are same, then it can forms two routes from R1 to R4, but the next hop is different. If two routes are selected as the best, then they form the equal-cost route.

4.8.2 ECMP Configuration Task List

1. Configure the max number of equal-cost route

1. Configure the max number of equal-cost route

Command	Explanation
Global mode	
maximum-paths <1-32> **no maximum-paths**	Configure the max number of equal-cost route.

4.8.3 ECMP Typical Example

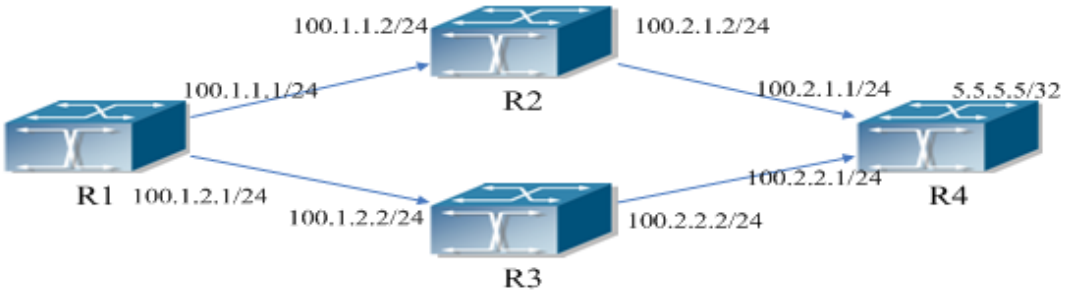


Figure 4-18 the application environment of ECMP

As it is shown in the figure, the R1 connect to R2 and R3 with the interface address 100.1.1.1/24 and 100.1.2.1/24. The R2 and R3 connect to R1 with the interface address 100.1.1.2/24 and 100.1.2.2/24. The R4 connect to R2 and R3 with interface address 100.2.1.1/24 and 100.2.2.1/24. The R2 and R3 connect to R4 with the interface address 100.2.1.2/24, 100.2.2.2/24. The loopback address of R4 is 5.5.5.5/32.

4.8.3.1 Static Route Implements ECMP

R1(config)#ip route 5.5.5.5/32 100.1.1.2
R1(config)#ip route 5.5.5.5/32 100.1.2.2
On R1, show ip route, the following is displayed:
R1(config)#show ip route
Codes: K - kernel, C - connected, S - static, R - RIP, B - BGP
O - OSPF, IA - OSPF inter area
N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
E1 - OSPF external type 1, E2 - OSPF external type 2
i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, ia - IS-IS inter area
* - candidate default
C 1.1.1.1/32 is directly connected, Loopback1 tag:0
S 5.5.5.5/32 [1/0] via 100.1.1.2, Vlan100 tag:0
[1/0] via 100.1.2.2, Vlan200 tag:0
C 100.1.1.0/24 is directly connected, Vlan100 tag:0
C 100.1.2.0/24 is directly connected, Vlan200 tag:0
C 127.0.0.0/8 is directly connected, Loopback tag:0
Total routes are : 6 item(s)

4.8.3.2 OSPF Implements ECMP

R1 configuration:
R1(config)#interface Vlan100
R1(Config-if-Vlan100)# ip address 100.1.1.1 255.255.255.0
R1(config)#interface Vlan200
R1(Config-if-Vlan200)# ip address 100.1.2.1 255.255.255.0
R1(config)#interface loopback 1
R1(Config-if-loopback1)# ip address 1.1.1.1 255.255.255.255
R1(config)#router ospf 1
R1(config-router)# ospf router-id 1.1.1.1
R1(config-router)# network 100.1.1.0/24 area 0
R1(config-router)# network 100.1.2.0/24 area 0

R2 configuration:
R2(config)#interface Vlan100
R2(Config-if-Vlan100)# ip address 100.1.1.2 255.255.255.0

```
R2(config)#interface Vlan200
R2(Config-if-Vlan200)# ip address 100.2.1.2 255.255.255.0
R2(config)#interface loopback 1
R2(Config-if-loopback1)# ip address 2.2.2.2 255.255.255.255
R2(config)#router ospf 1
R2(config-router)# ospf router-id 2.2.2.2
R2(config-router)# network 100.1.1.0/24 area 0
R2(config-router)# network 100.2.1.0/24 area 0
```

R3 configuration:

```
R3(config)#interface Vlan100
R3(Config-if-Vlan100)# ip address 100.1.2.2 255.255.255.0
R3(config)#interface Vlan200
R3(Config-if-Vlan200)# ip address 100.2.2.2 255.255.255.0
R3(config)#interface loopback 1
R3(Config-if-loopback1)# ip address 3.3.3.3 255.255.255.255
R3(config)#router ospf 1
R3(config-router)# ospf router-id 3.3.3.3
R3(config-router)# network 100.1.2.0/24 area 0
R3(config-router)# network 100.2.2.0/24 area 0
```

R4 configuration:

```
R4(config)#interface Vlan100
R4(Config-if-Vlan100)# ip address 100.2.1.1 255.255.255.0
R4(config)#interface Vlan200
R4(Config-if-Vlan200)# ip address 100.2.2.1 255.255.255.0
R4(config)#interface loopback 1
R4(Config-if-loopback1)# ip address 5.5.5.5 255.255.255.255
R4(config)#router ospf 1
R4(config-router)# ospf router-id 4.4.4.4
R4(config-router)# network 100.2.1.0/24 area 0
R4(config-router)# network 100.2.2.0/24 area 0
```

On R1, show ip route, the following is displayed:

```
R1(config)#show ip route
Codes: K - kernel, C - connected, S - static, R - RIP, B - BGP
       O - OSPF, IA - OSPF inter area
       N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
       E1 - OSPF external type 1, E2 - OSPF external type 2
       i - IS-IS, L1 - IS-IS level-1, L2 - IS-IS level-2, ia - IS-IS inter area
       * - candidate default
C        1.1.1.1/32 is directly connected, Loopback1    tag:0
O        5.5.5.5/32 [110/3] via 100.1.1.2, Vlan100, 00:00:05    tag:0
                    [110/3] via 100.1.2.2, Vlan200, 00:00:05    tag:0
```

C 100.1.1.0/24 is directly connected, Vlan100 tag:0
C 100.1.2.0/24 is directly connected, Vlan200 tag:0
O 100.2.1.0/24 [110/2] via 100.1.1.2, Vlan100, 00:02:25 tag:0
O 100.2.2.0/24 [110/2] via 100.1.2.2, Vlan200, 00:02:25 tag:0
C 127.0.0.0/8 is directly connected, Loopback tag:0
Total routes are : 8 item(s)

4.8.4 ECMP Troubleshooting

When configuring ECMP, ECMP may not run normally for the reasons of physical connection and false configuration, so users should note the following essential.

☞ When using ECMP, load-balance mode should be set as dst-src-ip or dst-src-mac-ip, after that, load-balance is correct for packets.

4.9 BFD

4.9.1 Introduction to BFD

BFD (Bidirectional Forwarding Detection) provides a detection mechanism to quickly detect and monitor the connectivity of links in networks. To improve network performance, between protocol neighbors must quickly detect communication failures to restore communication through backup paths as soon as possible.

BFD provides a general-purpose, standard, medium-independent and protocol-independent fast failure detection mechanism. It can uniformly and quickly detect the failures of the bidirectional forwarding paths between two network devices for superstratum protocols, such as routing protocols and Multiprotocol Label Switching (MPLS). BFD establishes session between two network devices to monitor their bidirectional forwarding paths to serve for superstratum protocols. However, there is no discovery mechanism for BFD, it is notified by superstratum protocol to establish sessions. After a session is established, if no BFD control packet is received from the peer within detection time, it notifies the failure to superstratum protocol which will take appropriate measures.

4.9.2 BFD Configuration Task List

1. Configure BFD basic function
2. Configure BFD for RIP (ng)
3. Configure BFD for static route (IPv6)
4. Configure BFD for VRRP (v3)

1. Configure BFD basic function

Command	Explanation
Global Mode	
bfd mode{active | passive} **no bfd mode**	Configure the mode before establishing BFD session, the default is active mode. No command restores active mode.
bfd authentication key <*1-255*> text <*WORD*> **no bfd authentication key <*1-255*>**	Configure key and authentication character string encrypted with text for BFD, no command deletes the configured key.
bfd authentication key <*1-255*> md5 <*WORD*> **no bfd authentication key**	Configure key and authentication character string encrypted with md5 for BFD, no command deletes the configured key.
Interface Mode	
bfd interval <*value1*> min_rx <*value2*> multiplier <*value3*> **no bfd interval**	Configure the minimum transmission interval and the multiplier of session detection for BFD control packets, no command restores the default detection multiplier.
bfd min-echo-receive-interval <*value*> **no bfd min-echo-receive-interval**	Configure the minimum receiving interval for BFD control packets, no command restores its default value.
bfd echo **no bfd echo**	Enable bfd echo, no command disables the function.
bfd echo-source-ip <*ipv4-address*> **no bfd echo-source-ip**	Detect link fault by configuring source address of echo packets, no command deletes the configured source address of echo packets.
bfd echo-source-ipv6 <*ipv6-address*> **no bfd echo-source-ipv6**	Detect link fault by configuring source address of echo packets, no command deletes the configured source address of echo packets.
bfd authentication key <*1-255*> **no bfd authentication key**	Enable BFD authentication and configure key for interface, no command disables BFD

	authentication.

2. Configure BFD for RIP (ng)

Command	Explanation
Interface Mode	
rip bfd enable **no rip bfd enable**	Configure BFD for RIP protocol on the specific interface, no command disables BFD for RIP protocol.
ipv6 rip bfd enable **no ipv6 rip bfd enable**	Configure BFD for RIPng protocol on the specific interface, no command cancels the configuration.

3. Configure BFD for static route (IPv6)

Command	Explanation
Global Mode	
ip route {vrf <name> <ipv4-address> | <ipv4-address>} mask <nexthop> bfd **no ip route {vrf <name> <ipv4-address> | <ipv4-address>} mask <nexthop> bfd**	Configure BFD for the static route, no command cancels the configuration.
ipv6 route {vrf <name> <ipv6-address> | <ipv6-address>} prefix <nexthop> bfd **no ipv6 route {vrf <name> <ipv6-address> | <ipv6-address>} prefix <nexthop> bfd**	Configure BFD for the static IPv6 route, no command cancels the configuration.

4. Configure BFD for VRRP (v3)

Command	Explanation
VRRP(v3) Group Configuration Mode	
bfd enable **no bfd enable**	Enable BFD for VRRP(v3) protocol and enable BFD detection on this group, no command disables the function.

4.9.3 Examples of BFD

4.9.3.1 Example for Linkage of BFD and Static Route

Example:

Configure a static route to 14.1.1.0/24 on Switch A and configure a static route to 15.1.1.0/24 on Switch B. Both switches enable BFD detection. When the link between Switch A and Switch B is failing, BFD can detect it immediately.



Figure 4-19

Configuration procedure:

Switch A:

Switch#config

Switch(config)#interface vlan 12

Switch(config-if-vlan12)#ip address 12.1.1.1 255.255.255.0

Switch(config)#interface vlan 15

Switch(config-if-vlan15)#ip address 15.1.1.1 255.255.255.0

Switch(config)#ip route 14.1.1.0 255.255.255.0 12.1.1.2 bfd

Switch B:

Switch#config

Switch(config)#interface vlan 12

Switch(config-if-vlan12)#ip address 12.1.1.2 255.255.255.0

Switch(config)#interface vlan 14

Switch(config-if-vlan15)#ip address 14.1.1.1 255.255.255.0

Switch(config)#ip route 15.1.1.0 255.255.255.0 12.1.1.1 bfd

When the link between Switch B and layer 2 switch is failing, Switch A can detect the change of Switch B immediately, here the static routing is at inactive state.

4.9.3.2 Example for Linkage of BFD and RIP Route

Example:

Switch A and Switch B are connected and run RIP protocol, both of them enable BFD function. When the link between Switch A and Switch B is failing, BFD can detect it immediately.

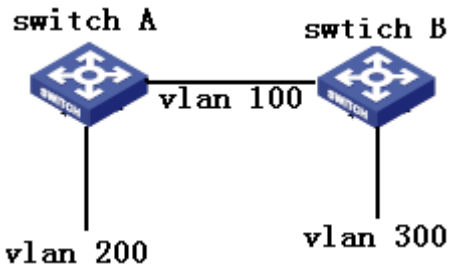


Figure 4-20

Configuration procedure:

Switch A:

Switch#config

Switch(config)#bfd mode active

Switch(config)#interface vlan 100

Switch(config-if-vlan100)#ip address 10.1.1.1 255.255.255.0

Switch(config)#interface vlan 200

Switch(config-if-vlan200)#ip address 20.1.1.1 255.255.255.0
Switch(config)#router rip
Switch (config-router)#network vlan 100
Switch (config-router)#network vlan 200
Switch(config)#interface vlan 100
Switch(config-if-vlan100) #rip bfd enable
Switch B:
Switch#config
Switch(config)#bfd mode passive
Switch(config)#interface vlan 100
Switch(config-if-vlan100)#ip address 10.1.1.2 255.255.255.0
Switch(config)#interface vlan 300
Switch(config-if-vlan300)#ip address 30.1.1.1 255.255.255.0
Switch(config)#router rip
Switch (config-router)#network vlan 100
Switch (config-router)#network vlan 300
Switch(config)#interface vlan 100
Switch(config-if-vlan100) #rip bfd enable
When the link between Switch A and Switch B is failing, BFD can detect it immediately and notifies RIP to delete the learnt route.

4.9.3.3 Example for Linkage of BFD and VRRP

Example:

When the master is failing, the backup cannot become the master until the configured timeout timer expires. The timeout is generally three to four seconds and therefore the switchover is slow. To solve this problem, VRRP uses BFD to probe the state of the master. Once the master fails, the backup can become the new master within 100 ms.

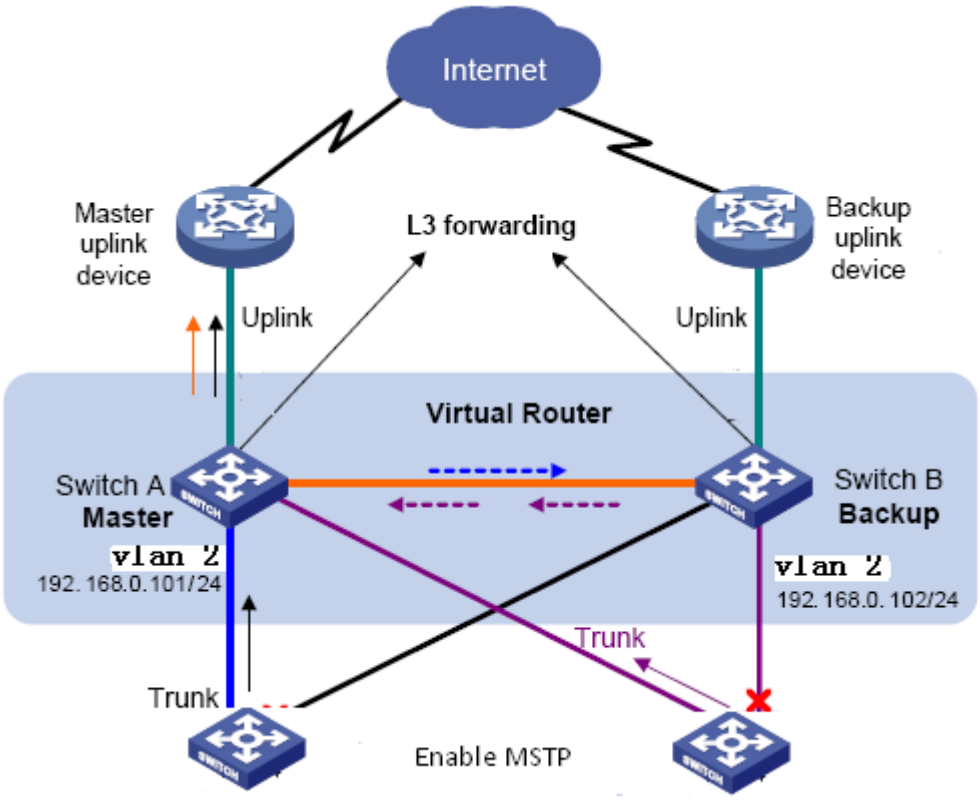


Figure 4-21

Configuration procedure:

Configure Switch A

Switch#config

Switch(config)#bfd mode active

Switch(config)#interface vlan 2

Switch(config-ip-vlan2)#ip address 192.16.0.101 255.255.255.0

Switch(config)#router vrrp 1

Switch(config-router)#virtual-ip 192.168.0.10

Switch(config-router)#interface vlan 1

Switch(config-router)#enable

Switch(config-router)#bfd enable

Configure Switch B

Switch#config

Switch(config)#bfd mode passive

Switch(config)#interface vlan 2

Switch(config-ip-vlan2)#ip address 192.16.0.102 255.255.255.0

Switch(config)#router vrrp 1

Switch(config-router)#virtual-ip 192.168.0.10

Switch(config-router)#interface vlan 1

Switch(config-router)#enable

Switch(config-router)#bfd enable

4.9.4 BFD Troubleshooting

When the problem of BFD function happens, please check whether the problem is resulted by the following reasons:

- Check whether the route protocol neighbor is established successfully. If no route protocol neighbor is established successfully, here BFD can not process the detection.
- Check whether the configured source-ip is correct for linkage with static route, if the connectivity of IP between two peers fails, BFD can not process the detection.
- Check whether VRRP group is established successfully for linkage with VRRP protocol. If no VRRP group is established successfully, here BFD can not process the detection.

4.10 BGP GR

4.10.1 Introduction to GR

Along with network development, it requires the higher availability, so HA (High Availability) is set, namely, how to ensure packets to be forwarded and does not affect traffic operation when router control layer can not work normally.

Usually, when a router does not work normally, neighbor in route protocol layer will detect their relationship to be down, and is up soon. The process is called neighborhood shock. This shock will result the router shock that will eventually result router black hole or data passed by restarted router. Finally network availability will decrease quickly.

In order to achieve high availability, it needs upper layer route protocol to support GR (Graceful Restart). Use GR can ensure that packets can be processed or forwarded correctly when the control layer is failing.

GR can reduce route shock, resource expend consumption of control layer and improve network stability. What describe in this document is GR, which can restart BGP protocol without affecting forwarding process, and forward packets in the correct path.

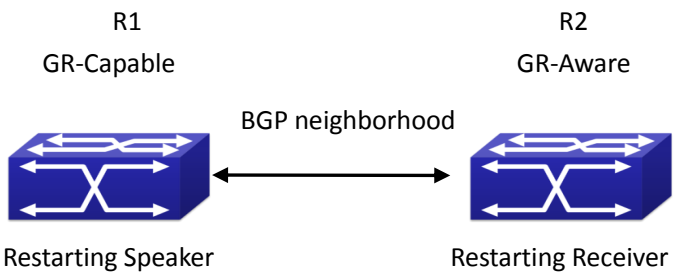


Figure 4-22 Aplication environment for GR

GR needs cooperation of GR-Capable router and GR-Aware router to complete. A restarted router is call Restarting Speaker（or GR-Restarter）, and its neighbor can be called Receiving Speaker（or GR-Helper）. Restarting Speaker is GR-Capable router while Receiving Speaker is GR-Aware router. In this way, they can complete GR. Suppose that router R1 and R2 establish BGP neighborhood, as shown in Fig 13-1, GR process can be described as:

Restarting Speaker（GR-Restarter）:

1. R1 and R2 negotiate GR capability through OPEN when establish original BGP neighbor.

2. When R1 is restarted, the route is kept in the interface board and guide forwarding continuously.

3. R1 establishes TCP connection with R2 again, it sets Restart state to 1 in BGP OPEN messages to show that this router has been restarted. At the same time, it will inform the value of restart time (it is less than Holdtime in OPEN messages) to neighbor. Additionally, it should inform neighbor what type of GR is supported.

4. After R1 is established connection with R2 correctly, it is able to receive and deal with the update information and enable selection deferral timer.

5. R1 delays the count process of the local BGP route until it receives all End-of-RIB from BGP neighbors in GR-Aware or until the local selection deferral timer is overtime.

6. Count route and send the update route. After that, it will send End-of-RIB to neighbors.

Restarting Speaker（GR-Helper）:

1. R1 and R2 negotiate GR capability with the restarted router when they establish the original neighborhood with BGP, R1 is a router that support GR-Capable.

2. When R1 is restarted, R2 may senses that TCP between R1 and R2 is cut off or cannot detect the previous state before they establish TCP connection again. If it does not detect it, go to step 4, otherwise go to step 3.

3. Keep the route sent by R1 and mark a stale label. After that, enable Restart Timer.

4. Cut off old TCP connection and deal with new TCP connection continuously. Keep the route sent by R1 and mark a stale label. After that, enable Restart Timer.

5. Establish a new neighborhood with the restarted router, delete Restart Timer and enable Stale Path Timer.

6. Before establish the new neighborhood, If Restart Timer is overtime, Restart flag does not equal 1, or there is no relevant supporting information in AFI/SAFI address family, please clear the kept route.

7. Send the route update information to the restarted router, after that, it will send End-Of-RIB label.

8. If Stale Path Timer is overtime, clear the kept route.

4.10.2 GR Configuration Task List

1. Configure whether GR capablility is supported
2. Configure whether the specific neighbor supports GR capablility
3. Configure restart-time
4. Configure restart-time for neighbor

5. Configure stale-path-time for BGP GR
6. Configure selection-deferral-time for BGP GR

1. Configure whether GR capablility is supported

Command	Description
BGP route configuration mode	
bgp graceful-restart **no bgp graceful-restart**	Enable BGP to support GR.

2. Configure whether the specific neighbor supports GR capablility

Command	Description
BGP protocol unicast address family mode and VRF address family mode	
neighbor (A.B.C.D | X:X::X:X | WORD) capability graceful-restart **no neighbor (A.B.C.D | X:X::X:X | WORD) capability graceful-restart**	Set a label for neighbor, it takes GR parameter when send OPEN messages.

3. Configure restart-time

Command	Description
BGP route configuration mode	
bgp graceful-restart restart-time <1-3600> **no bgp graceful-restart restart-time <1-3600>**	Configure BGP GR's restart-time (Receiving Speaker enables a timeout timer for a neighbor, it uses the restart-time as the timeout). A restart-time specifies the longest waiting time from Receiving Speaker finds restarting to the received OPEN messages. If Receiving Speaker does not receive OPEN messages after exceed the time, it can delete SATLE route saved by neighbor.

4. Configure restart-time for neighbor

Command	Discription
BGP protocol unicast address family mode and VRF address family mode	
neighbor (A.B.C.D | X:X::X:X | WORD) restart-time <1-3600> **no neighbor (A.B.C.D | X:X::X:X | WORD) restart-time <1-3600>**	Configure restart-time for neighbors, no command restores the default time.

5. Configure stale-path-time for BGP GR

Command	Discription
BGP route configuration mode	

bgp graceful-restart stale-path-time <1-3600> **no bgp graceful-restart stale-path-time <1-3600>**	Stalepath-time uses the default value of 360s, which is much longer than restart-time and selection-deferral-time. Because during the time from Receiving Speaker receives OPEN messages to receives EOR, it sends the initial route update and waits that the initial route update is received completely.

6. Configure selection-deferral-time for BGP GR

Command	Description
BGP route configuration mode	
bgp selection-deferral-time <1-3600> **no bgp selection-deferral-time <1-3600>**	Specify the longest waiting time that start to count selection route from the received OPEN messages to the received EOR for Restarting Speaker. If Restarting Speaker does not receive EOR after exceed the time, it can count selection route.

4.10.3 Typical Example of GR

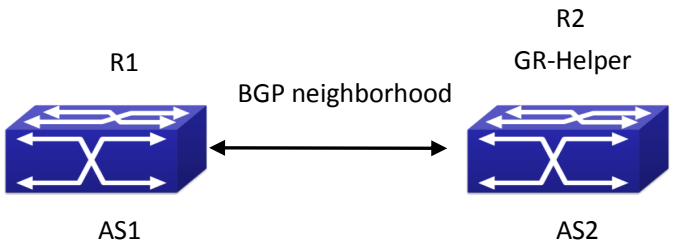


Figure 4-23 Application environment for GR

As shown in Figure 4-23, R1 and R2 establish BGP neighborhood. When they cut off the connection, BGP protocol in R2 will be in helper mode, keep route entry from R1 to R2 and restart restart-time timer. During that time, if it receives open messages from R1 or timer is overtime, the route marked with stale in R2 will be deleted. When they establish neighborhood again, R1 will enable selection timer and wait R2 to send EOR messages or until timer is overtime, R1 is a selection route. However, after R2 receives open messages from R1, it enables STALE route timer and it will delete timer and stale route when accept EOR from R1 or timer is overtime.

R1 configures int vlan 12， ip address 12.1.1.1
R2 configuresint vlan 12， ip address 12.1.1.2
R1 configuration:
R1#config
R1(config)#vlan 12
R1(config-vlan12)#int vlan 12

```
R1(config-if-vlan12)#ip address 12.1.1.1 255.255.255.0
R1(config-if-vlan12)#exit
R1(config)#router bgp 1
R1(config-router)#neighbor 12.1.1.2 remote-as 2
R1(config-router)#neighbor 12.1.1.2 capability graceful-restart
R1(config-router)#bgp selection-deferral-time 120
R1(config-router)#bgp graceful-restart restart-time 60
R1(config-router)#bgp graceful-restart stale-path-time 180
R1(config-router)#exit
```

R2 configuration:

```
R2#config
R2(config)#vlan 12
R2(config-vlan12)#int vlan 12
R2(config-if-vlan12)#ip address 12.1.1.2 255.255.255.0
R2(config-if-vlan12)#exit
R2(config)#router bgp 2
R2(config-router)#neighbor 12.1.1.1 remote-as 1
R2(config-router)#neighbor 12.1.1.1 capability graceful-restart
R2(config-router)#bgp selection-deferral-time 120
R2(config-router)#bgp graceful-restart restart-time 60
R2(config-router)#bgp graceful-restart stale-path-time 180
R2(config-router)#exit
```

4.11 OSPF GR

4.11.1 Introduction to OSPF GR

OSPF Graceful-Restart（short for OSPF GR），is used to maintain data forwarding correctly and flow of crucial service is not interrupted when routing protocol restarts or switchover of layer 3 switches between active master and standby master. It is one of high availability technologies.

So far, the high layer 3 switches usually adopt a design for separating control and forwarding. The control module for counting routing protocol at master control board, but data forwarding module is at liner card. As a result, it will not affect data forwarding on line card when the master control board is restarted. So the device supporting GR is generally a chassis device and has two master control boards.

Since standard OSPF protocol (RFC2328) does not support GR, it will lead to flow cut off and routing surge when routing protocol is restarted or switchover between active master and standby master for various reasons. For example, as shown in below figure, when S1 occurs switchover, the neighborhood relation between S1 and S2 will lose, at that time S2 will send Router-LSA to S3 and S4 and this LSA does not include the link between S1 and S2. After S3 and S4 received LSA, they will count routing protocol again. The result will not include the link

between S1 and S2. After S1 finishes the switchover, it will establish neighborhood relation with S2 and synchronize database, this action leads S2, S3 and S4 to count routing again. However, switchover of S1 will result routing shiver, which is not accepted by some networks with high requirement for performance.

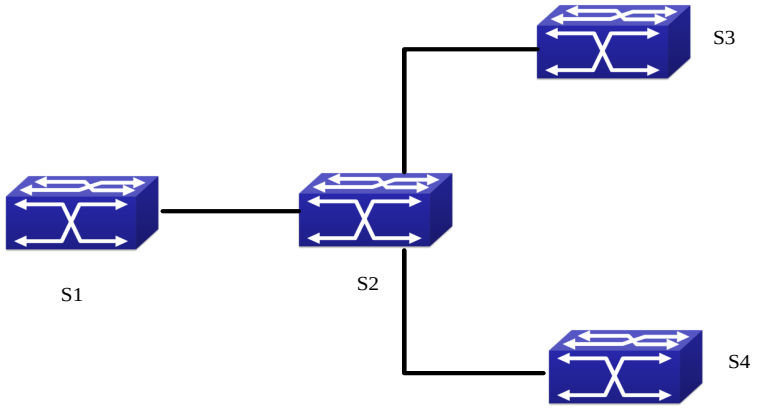


Figure 4-24 typical application scene

OSPF GR described in RF C3623 is come up for the above state. Its basic idea is that if the network topology keeps stabilization during the switchover and layer 3 switch can maintain the same forwarding list, then its neighbor can maintain their relationship, which can make the switch on its forwarding path still. If S1 and S2 support and enable GR, the liner card of S1 will keep the traffic forwarding and S2 can maintain the relationship with S1, at the same time, network topology between S3 and S4 will not be changed, furthermore, it does not need to count routing again. All of these ensure the traffic forwarding and avoid routing shiver.

Layer 3 switch can be divided into GR restarter and GR helper according to its function in GR process. GR restarter is layer 3 switch to occur the switchover between active master and standby master or restart protocol while GR helper is layer 3 switch to help GR restarter. In the above example, S1 is GR restarter and S2 is GR helper

The advantages of OSPF GR in the following:

- Increase network reliability
- Reduce the effect of routing shiver to network
- Reduce the effect to traffic and avoid that lose packets during switchover

4.11.2 OSPF GR Configuration

OSPF GR configuration task list:

1. Enable GR for OSPF
2. Configure grace-period for OSPF GR restarter (optional)
3. Configure policy for OSPF GR helper (optional)

1. Enable GR for OSPF

Command	Description
OSPF protocol configuration mode	

capability restart graceful **no capability restart**	Enable GR of specific OSPF.

2. Configure grace-period for OSPF GR restarter (optional)

Command	Description
Global configuration mode	
ospf graceful-restart grace-period <*integer*> **no ospf restart grace-period**	Configure grace period for GR restarte (The switch is used to the switchover or restart the protocol). The no command restores its default value.

3. Configure the policy for OSPF GR helper (optional)

Command	Description
Global configuration mode	
ospf graceful-restart helper max-grace-period <*integer*> **no ospf graceful-restart helper**	One of GR helper policy. Configure maximum grace period supported by helper. The no command deletes all configured helper policy.
ospf graceful-restart helper never **no ospf graceful-restart helper**	One of GR helper policy. Configure the switch can not become OSPF GR helper. The no command deletes all configured helper policy.

4.11.3 OSPF GR Example

Example:

There are for switches from S1 to S4 (They are two master control board and supports OSPF GR), they enable OSPF to implement the following functions:

1. S1 keeps traffic forwarding during the switchover, S2-S4 ensure that no routing shiver and the continuous network traffic.
2. S1 needs to finish the switchover and restart protocol within 120s, otherwise S2 will quit GR and count routing again.
3. S1 does not work as a OSPF GR Helper (S1 will not help S2 to process GR, but it will count routing again when S2 processes the switchover or restart OSPF protocol).

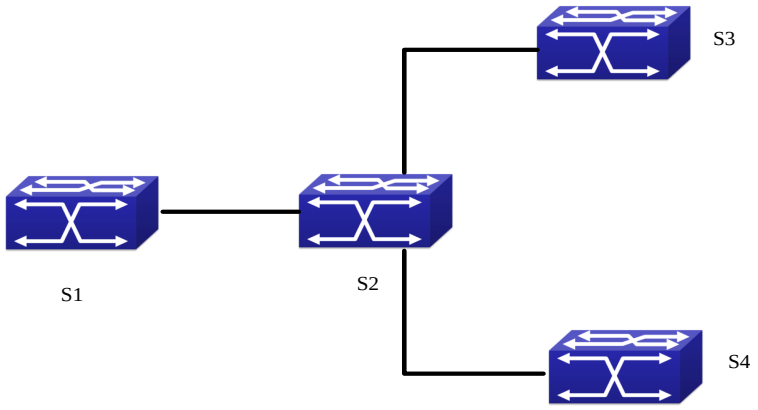


Figure 4-25 Typical application

Steps: Since the switch enables OSPF GR by default, we only need to configure the parameters and helper policy for OSPF GR. (the following configuration is relative with OSPF GR only and that of topology is omitted).

S1

S1(config)#ospf graceful-restart grace-period 120

S1(config)# ospf graceful-restart helper never

S2

S2(config)# ospf graceful-restart helper max-grace-period 120

4.11.4 OSPF GR Troubleshooting

When you have trouble in using OSPF GR, please check the following reasons:

- Whether GR restarter switch supports OSPF GR and has two main control boards, please ensure that specific GR is not disabled.
- Whether network topology is changed during OSPF GR process. When it is changed, switch may quit GR and restart OSPF.
- Please ensure all neighbors of GR restarter support GR.
- Do not modify the relevant configuration of OSPF during GR.

Chapter 5 Multicast Protocol Related Configuration

5.1 Multicast

This chapter will give an introduction to the configuration of IPv4 Multicast Protocol.

5.1.1 Introduction to Multicast

Various transmission modes can be adopted when the destination of packet (including data, sound and video) transmission is the minority users in the network. One way is to use Unicast mode, i.e. to set up a separate data transmission path for each user; or, to use Broadcast mode, which is to send messages to all users in the network, and they will receive the Broadcast messages no matter they need or not. For example, if there are 200 users in a network who want to receive the same packet, then the traditional solution is to send this packet for 200 times separately via Unicast to guarantee the users who need the data can get all data wanted, or send the data in the entire domain via Broadcast. Transferring the data in the whole range of network .The users who need these data can get directly from the network. Both modes waste a great deal of valuable bandwidth resource, and furthermore, Broadcast mode goes against the security and secrecy.

The emergence of IP Multicast technology solved this problem in time. The Multicast source only sends out the message once, Multicast Routing Protocol sets up tree-routing for Multicast data packet, and then the transferred packet just starts to be duplicated and distributed in the bifurcate crossing as far as possible. Thus the packet can be sent to every user who needs it accurately and effectively.

It should be noticed that it is not necessary for Multicast source to join in Multicast group. It sends data to some Multicast groups, but it is not necessarily a receiver of the group itself. There can be more than one source sending packets to a Multicast group simultaneously. There may exist routers in the network which do not support Multicast, but a Multicast router can encapsulate the Multicast packets into Unicast IP packets with tunnel mode to send them to the Multicast router next to it, which will take off the Unicast IP header and continue the Multicast transmission process, thus a big alteration of network structure is avoided. The primary advantages of Multicast are:

1. Enhance efficiency: reduce network traffic, lighten the load of server and CPU
1. Optimize performance: reduce redundant traffic
1. Distributed application: Enable Multipoint Application

5.1.2 Multicast Address

The destination address of Multicast message uses class D IP address with range from 224.0.0.0 to 239.255.255.255. D class address can not appear in the source IP address field of an IP message. In the process of Unicast data transmission, the transmission path of a data packet is from source address routing to destination address, and the transmission is performed with hop-by-hop principle. However, in IP Multicast environment, the destination addresses is a group instead of a single one, they form a group address. All message receivers will join in a group, and once they do, the data flowing to the group address will be sent to the receivers immediately and all members in the group will receive the data packets. The members in a Multicast group are dynamic, the hosts can join and leave the Multicast group at any time.

Multicast group can be permanent or temporary. Some of the Multicast group addresses are assigned officially; they are called Permanent Multicast Group. Permanent Multicast Group keeps its IP address fixed but its member structure can vary within. The member amount of Permanent Multicast Group can be arbitrary, even zero. The IP Multicast addresses which are not kept for use by Permanent Multicast Group can be utilized by temporary Multicast groups.

224.0.0.0～224.0.0.255 are reserved Multicast addresses (Permanent Group Address), address 224.0.0.0 is reserved but not assigned, and other addresses are used by Routing Protocol; 224.0.1.0～238.255.255.255 are Multicast addresses available to users（Temporary Group Address） and are valid in the entire domain of the network; 239.0.0.0～239.255.255.255 are local management Multicast addresses, which are valid only in specific local domain. Frequently used reserved multicast address list is as follows:

Benchmark address (reserved)
224.0.0.1 Address of all hosts
224.0.0.2 Address of all Multicast Routers
224.0.0.3 Unassigned
224.0.0.4 DVMRP Router
224.0.0.5 OSPF Router
224.0.0.6 OSPF DR
224.0.0.7 ST Router
224.0.0.8 ST host
224.0.0.9 RIP-2 Router
224.0.0.10 IGRP Router
224.0.0.11 Active Agent
224.0.0.12 DHCP Server/Relay Agent
224.0.0.13 All PIM Routers
224.0.0.14 RSVP Encapsulation
224.0.0.15 All CBT Routers
224.0.0.16 Specified SBM
224.0.0.17 All SBMS
224.0.0.18 VRRP
224.0.0.22 IGMP

When Ethernet transmits Unicast IP messages, the destination MAC address it uses is the receiver's MAC address. But in transmitting Multicast packets, the transmission destination is not a specific receiver any more, but a group with uncertain members, thus Multicast MAC address is used. Multicast MAC address is corresponding to Multicast IP address. It is prescribed in IANA

(Internet Assigned Number Authority) that the higher 25 bits in Multicast MAC address is 0x01005e, and the lower 23bits in MAC address is the lower 23bits in Multicast IP address.

Since only 23bits out of the lower 28bits in IP Multicast address are mapped into MAC address, therefore there are 32 IP Multicast addresses which are mapped into the same MAC address.

5.1.3 IP Multicast Packet Transmission

In Multicast mode, the source host sends packets to the host group indicated by the Multicast group address in the destination address field of IP data packet. Unlike Unicast mode, Multicast data packet must be forwarded to a number of external interfaces to be sent to all receiver sites in Multicast mode, thus Multicast transmission procedure is more complicated than Unicast transmission procedure.

In order to guarantee that all Multicast packets get to the router via the shortest path, the receipt interface of the Multicast packet must be checked in some certain way based on Unicast router table; this checking mechanism is the basis for most Multicast Routing Protocol to forward in Multicast mode --- RPF (Reverse Path Forwarding) check. Multicast router makes use of the impressed packet source address to query Unicast Router Table or independent Multicast Router Table to determine if the packet ingress interface is on the shortest path from receipt site to source address. If shortest path Tree is used, then the source address is the address of source host which sends Multicast Data Packets; if Shared Tree is used, then the source address is the address of the root of the Shared-Tree. When Multicast data packet gets to the router, if RPF check passes, then the data packet is forwarded according to Multicast forward item, and the data packet will be discarded else wise.

5.1.4 IP Multicast Application

IP Multicast technology has effectively solved the problem of sending in single point and receiving in multipoint. It has achieved the effective data transmission from a point to multiple points, saved a great deal of network bandwidth and reduced network load. Making use of the Multicast property of network, some new value-added operations can be supplied conveniently. In Information Service areas such as online living broadcast, network TV, remote education, remote medicine, real time video/audio meeting, the following applications may be supplied:

1) Application of Multimedia and Streaming Media
2) Data repository, finance application (stock) etc
3) Any data distribution application of “one point to multiple points”

In the situation of more and more multimedia operations in IP network, Multicast has tremendous market potential and Multicast operation will be generalized and popularized.

5.2 PIM-DM

5.2.1 Introduction to PIM-DM

PIM-DM（Protocol Independent Multicast, Dense Mode）is a Multicast Routing Protocol in dense mode which applies to small network. The members of multicast group are relatively dense under this kind of network environment.

The working process of PIM-DM can be summarized as: Neighbor Discovery, Flooding & Prune, and Graft.

1. Neigh hour Discovery

After PIM-DM router is enabled, Hello message is required to discover neighbors. The network nodes which run PIM-DM use Hello message to contact each other. PIM-DM Hello message is sent periodically.

2. Flooding & Prune of process

PIM-DM assumes all hosts on the network are ready to receive Multicast data. When some Multicast Source begins to send data to a Multicast Group G, after receiving the Multicast packet, the router will make RPF check first according to the Unicast table. If the check passes, the router will create a (S, G) table entry and transmit the Multicast packet to all downstream PIM-DM nodes on the network (Flooding). If the RPF check fails, i.e. the Multicast packet is input from the incorrect interface, and then the message is discarded. After this procedure, in the PIM-DM Multicast domain, every node will create a (S, G) table entry. If there is no Multicast group member in the downstream nodes, then a Prune message is sent to upstream nodes to notify them not to transmit data of this Multicast group any more. After receiving Prune message, the upstream nodes will delete the corresponding interface from the output interface list to which their Multicast transmission table entry (S, G) corresponds. Thus a SPT（Shortest Path Tree, SPT）tree with source S as root is created. The Prune process is initiated by leaf router first.

The process above is called Flooding & Prune process. Each pruned node also provides time-out mechanics at the same time. When Prune is timed-out, the router will restart Flooding & Prune process. The PIM-DM Flooding & Prune is periodically processed.

3. RPF Check

With RPF Check, PIM-DM makes use of existing Unicast routing table to establish a Multicast transmission tree initiating from data source. When a Multicast packet arrives, the router will determine whether the coming path is correct first. If the arrival interface is the interface connected to Multicast source indicated by Unicast routing, then this Multicast packet is considered to be from the correct path. Otherwise the Multicast packet is to be discarded as redundant message. The Unicast routing message used as path judgment can root in any Unicast Routing Protocol, such as messages found by RIP, OSPF, etc. It doesn’t rely on any specific Unicast Routing Protocol.

4. Assert Mechanism

If each of two Multicast routers A and B on the same LAN segment has a receiving route respectively and both will transmit the Multicast packet to the LAN after receiving the Multicast data packet sent by the Multicast Source S, then the downstream node Multicast router C will receive two exactly same Multicast packets. The router needs to choose a unique transmitter

through Assert mechanism after it detects this situation. An optimal transmission path is selected through sending out Assert packet. If the priority and cost of two or more path are same, then the node with larger IP address is taken as the upstream neighbor of the (S, G) entry and in charge of the transmission of the (S, G) Multicast packet.

5. Graft

When the pruned downstream node needs to recover to transmission status, this node uses Graft Packet to notify upstream nodes to restore multicast data transmission.

5.2.2 PIM-DM Configuration Task List

1. Enable PIM-DM (Required)
2. Configure static multicast routing entries(Optional)
3. Configure additional PIM-DM parameters(Optional)
   a) Configure the interval for PIM-DM hello messages
   b) Configure the interval for state-refresh messages
   c) Configure the boundary interfaces
   d) Configure the management boundary
4. Disable PIM-DM protocol

1. Enable the PIM-DM protocol

When configuring the PIM-DM protocol on Layer 3 switches, PIM multicasting should be enabled globally, then PIM-DM can be enabled for specific interfaces.

Command	Explanation
Global Mode	
ip pim multicast-routing **no ip pim multicast-routing**	To enable PIM-DM globally for all the interfaces (However, in order to make PIM-DM work for specific interfaces, the following command should be issued).

And then turn on PIM-SM switch on the interface

Command	Explanation
Interface Configuration Mode	
ip pim dense-mode	To enable PIM-DM protocol for the specified interface.(Required)

2. Configure static multicast routing entries

Command	Explanation
Global Configuration Mode	
ip mroute <*A.B.C.D*> <*A.B.C.D*> <*ifname*> <*.ifname*> **no ip mroute <*A.B.C.D*> <*A.B.C.D*> [<*ifname*> <*.ifname*>]**	To configure a static multicast routing entry. The no form of this command will remove the specified entry.

3. Configure additional PIM-DM parameters

a) Configure the interval for PIM-DM hello messages

Command	Explanation
Interface Configuration Mode	
ip pim hello-interval < ***interval*****>** **no ip pim hello-interval**	To configure the interval for PIM-DM hello messages. The no form of this command will restore the interval to the default value.

b) Configure the interval for state-refresh messages

Command	Explanation
Interface Configuration Mode	
ip pim state-refresh origination-interval **no ip pim state-refresh origination-interval**	To configure the interval for sending PIM-DM state-refresh packets. The no form of this command will restore the default value.

c) Configure the boundary interfaces

Command	Explanation
Interface Configuration Mode	
ip pim bsr-border **no ip pim bsr-border**	To configure the interface as the boundary of PIM-DM protocol. On the boundary interface, BSR messages will not be sent or received. The network connected the interface is considered as directly connected network. The no form of this command will remove the configuration.

d) Configure the management boundary

Command	Explanation
Interface Configuration Mode	
ip pim scope-border <*1-99* **>|<***acl_name***>** **no ip pim scope-border**	To configure PIM-DM management boundary for the interface and apply ACL for the management boundary. With default settings, 239.0.0.0/8 is considered as the scope of the management group. If ACL is configured, then the scope specified by ACL permit command is the scope of the management group. The no form of this command will remove the configuration.

4. Disable PIM-DM protocol

Command	Explanation
Interface Configuration Mode	
no ip pim dense-mode	To disable the PIM-DM protocol for the interface.
Global Configuration Mode	
no ip pim multicast-routing	To disable PIM-DM globally.

5.2.3 PIM-DM Configuration Examples

As shown in the following figure, add the Ethernet interfaces of Switch A and Switch B to corresponding vlan, and enable PIM-DM Protocol on each vlan interface.

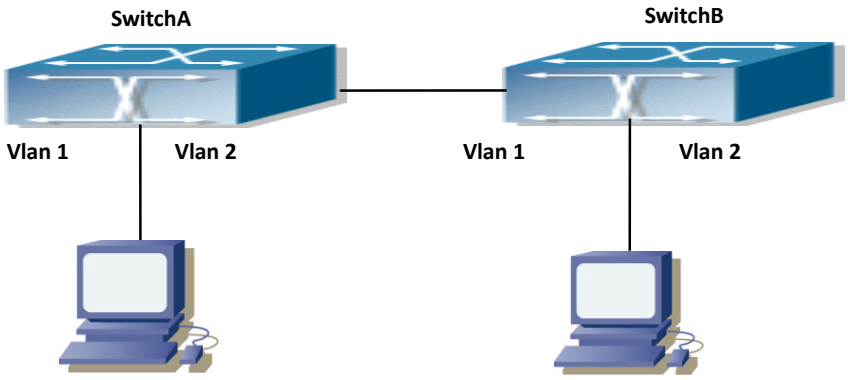


Figure 5-1 PIM-DM Typical Environment

The configuration procedure for SwitchA and SwitchB is as follows:

(1) Configure SwitchA:

Switch(config)#ip pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)# ip address 10.1.1.1 255.255.255.0
Switch(Config-if-Vlan1)# ip pim dense-mode
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan2
Switch(Config-if-Vlan2)# ip address 12.1.1.1 255.255.255.0
Switch(Config-if-Vlan2)# ip pim dense-mode

(2) Configure SwitchB:

Switch(config)#ip pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)# ip address 12.1.1.2 255.255.255.0
Switch(Config-if-Vlan1)# ip pim dense-mode
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)# ip address 20.1.1.1 255.255.255.0
Switch(Config-if-Vlan2)# ip pim dense-mode

At the same time, you should pay attention to the configuration of Unicast Routing Protocol, assure that each device can communicate with each other in the network layer, and be able to implement dynamic routing update in virtue of Unicast Routing Protocol.

5.2.4 PIM-DM Troubleshooting

In configuring and using PIM-DM Protocol, PIM-DM Protocol might not operate normally caused by physical connection or incorrect configuration. Therefore, the user should pay attention to the following issues:

- ☞ To assure that physical connection is correct
- ☞ To assure the Protocol of Interface and Link is UP (use show interface command)
- ☞ To assure PIM Protocol is enabled in Global Mode (use ipv6 pim multicast-routing)
- ☞ Enable PIM-DM Protocol on the interface (use ipv6 pim dense-mode command)
- ☞ Multicast Protocol requires RPF Check using Unicast routing; therefore the correctness of Unicast routing must be assured beforehand

If all attempts including Check are made but the problems on PIM-DM can't be solved yet, then use debug commands such as debug pim please, and then copy DEBUG information in 3 minutes and send to Technology Service Center.

5.3 PIM-SM

5.3.1 Introduction to PIM-SM

PIM-SM（Protocol Independent Multicast, Sparse Mode）is Protocol Independent Multicast Sparse Mode. It is a Multicast Routing Protocol in Sparse Mode and mainly used in big scale network with group members distributed relatively sparse and wide-spread. Unlike the Flooding & Prune of Dense Mode, PIM-SM Protocol assumes no host needs receiving Multicast data packets. PIM-SM router transmits Multicast Data Packets to a host only if it presents explicit requirement.

By setting RP (Rendezvous Point) and BSR (Bootstrap Router), PIM-SM announce Multicast packet to all PIM-SM routers and establish RPT (RP-rooted shared tree) based on RP using Join/Prune message of routers. Consequently the network bandwidth occupied by data packets and message control is cut down and the transaction cost of routers decreases. Multicast data get to the network segment where the Multicast group members are located along the shared tree flow. When the data traffic reaches a certain amount, Multicast data stream can be switched to the shortest path tree SPT based on the source to reduce network delay. PIM-SM doesn't rely on any specific Unicast Routing Protocol but make RPF Check using existing Unicast routing table.

1. PIM-SM Working Principle

The central working processes of PIM-SM are: Neighbor Discovery, Generation of RP Shared Tree (RPT), Multicast source registration, SPT Switch, etc. We won't describe the mechanism of Neighbor Discovery here since it is same as that of PIM-DM.

(1) Generation of RP Shared Tree (RPT)

When a host joins a Multicast Group G, the leaf router that is connected to this host directly finds out through IGMP message that there is a receiver of Multicast Group G, then it works out the corresponding Rendezvous Point RP for Multicast Group G, and send join message to upper lever nodes in RP direction. Every router on the way from the leaf router to RP will generate a (*, G) table entry, where a message from any source to Multicast group applies to this entry. When RP receives the message sent to Multicast Group G, the message will get to the leaf router along the set up path and reach the host. In this way the RPT with RP as root is generated.

(2) Multicast Source Registration

When a Multicast Source S sends a Multicast packet to Multicast Group G, the PIM-SM Multicast router connected to it directly will take charge of encapsulating the Multicast packet into registered message and unicast it to corresponding RP. If there are more than one PIM-SM Multicast routers on a network segment, then DR (Designated Router) takes charge of sending the Multicast packet.

(3) SPT Switch

When the Multicast router finds that the rate of the Multicast packet from RP with destination address G exceeds threshold, the Multicast router will send Join message to the next upper lever nodes in the source direction, which results in the switch from RPT to SPT.

2. Preparation before PIM-SM configuration

(1) Configuration Candidate RP

More than one RPs (candidate RP) can exist in PIM-SM network and each C-RP (Candidate RP) takes charge of transmitting Multicast packets with destination address in a certain range. To configure more than one candidate RPs can implement RP load share. No master or slave is differentiated among RPs. All Multicast routers work out the RP corresponding to some Multicast group based on the same algorithm after receiving the candidate RP message announced by BSR.

Note that one RP can serve more than one Multicast groups and all Multicast groups. Each Multicast group can only correspond to one unique RP at any moment. It can't correspond to more than one RP at the same time.

(2) Configure BSR

BSR is the management center of PIMSM network. It is in charge of collecting messages sent by candidate RPs and broadcast them.

Only one BSR can exist within a network, but more than one C-BSR (Candidate-BSR) can be configured. In this way, if some BSR goes wrong, it can switch to another. C-BSRs elect BSR automatically.

5.3.2 PIM-SM Configuration Task List

1. Enable PIM-SM (Required)
2. Configure static multicast routing entries (Optional)
3. Configure additional parameters for PIM-SM (Optional)
   (1) Configure parameters for PIM-SM interfaces
      1) Configure the interval for PIM-SM hello messages
      2) Configure the hold time for PIM-SM hello messages
      3) Configure ACL for PIM-SM neighbors
      4) Configure the interface as the boundary interface of the PIM-SM protocol
      5) Configure the interface as the management boundary of the PIM-SM protocol
   (2) Configure global PIM-SM parameters
      1)Configure the switch as a candidate BSR
      2)Configure the switch as a candidate RP

3)Configure static RP

4)Configure the cache time of kernel multicast route

4. Disable PIM-SM Protocol

1. Enable PIM-SM Protocol

The PIM-SM protocol can be enabled on Layer 3 switches by enabling PIM in global configuration mode and then enabling PIM-SM for specific interfaces in the interface configuration mode.

Command	Explanation
Global Mode	
ip pim multicast-routing	To enable the PIM-SM protocol for all the interfaces (However, in order to make PIM-SM work for specific interfaces, the following command should be issued).(Required)

And then turn on PIM-SM switch on the interface

Command	Explanation
Interface Configuration Mode	
ip pim sparse-mode	Enable PIM-SM Protocol of the interface. (Required).

2. Configure static multicast routing entries

Command	Explanation
Global Configuration Mode	
ip mroute <*A.B.C.D*> <*A.B.C.D*> <*ifname*> <*.ifname*> **no ip mroute <*A.B.C.D*> <*A.B.C.D*> [<*ifname*> <*.ifname*>]**	To configure a static multicast routing entry. The no form of this command will remove the specified static multicast routing entry.

3. Configure additional parameters for PIM-SM

(1) Configure parameters for PIM-SM interfaces

1)Configure the interval for PIM-SM hello messages

Command	Explanation
Interface Configuration Mode	
ip pim hello-interval <*interval*> **no ip pim hello-interval**	To configure the interval for PIM-SM hello messages. The no form of this command restores the interval to the default value.

2)Configure the hold time for PIM-SM hello messages

Command	Explanation
Interface Configuration Mode	
ip pim hello-holdtime <*value*> **no ip pim hello-holdtime**	To configure the value of the holdtime field in the PIM-SM hello messages. The no form of this command will restore the hold time to the default value.

3)Configure ACL for PIM-SM neighbors

Command	Explanation
Interface Configuration Mode	
ip pim neighbor-filter{<*access-list-number*> } **no ip pim neighbor-filter{<*access-list-number*> }**	To configure ACL to filter PIM-SM neighbors. If session to the neighbor has been denied by ACL, then the sessions that have been set up will be discarded immediately and new sessions will not be set up.

4)Configure the interface as the boundary interface of the PIM-SM protocol

Command	Explanation
Interface Configuration Mode	
ip pim bsr-border **no ip pim bsr-border**	To configure the interface as the boundary of PIM-SM protocol. On the boundary interface, BSR messages will not be sent or received. The network connected the interface is considered as directly connected network. The no form of this command will remove the configuration.

5)Configure the interface as the management boundary of the PIM-SM protocol

Command	Explanation
Interface Configuration Mode	
**ip pim scope-border <*1-99* >	<*acl_name*>** **no ip pim scope-border**	To configure PIM-SM management boundary for the interface and apply ACL for the management boundary. With default settings, 239.0.0.0/8 is considered as the scope of the management group. If ACL is configured, then the scope specified by ACL permit command is the scope of the management group. acl_name should be standard IPv4 ACL name. The no form of this command will remove the configuration.

(2) Configure global PIM-SM parameter

1) Configure the switch as a candidate BSR

Command	Explanation
Global Configuration Mode	
ip pim bsr-candidate {vlan <*vlan-id*>| <*ifname*>}[<*mask-length*>][<*priority*>] **no ip pim bsr-candidate**	This command is the global candidate BSR configuration command, which is used to configure the information of PIM-SM candidate BSR so that it can compete for BSR router with other candidate BSR. The "**no ip pim bsr-candidate**" command cancels the configuration of BSR.

2) Configure the switch as a candidate RP

Command	Explanation

Global Configuration Mode	
ip pim rp-candidate { vlan <*vlan-id*>| lookback<*index*> <*ifname*>} [<*A.B.C.D*>][<*priority*>] **no ip pim rp-candiate**	This command is the global candidate RP configuration command, which is used to configure the information of PIM-SM candidate RP so that it can compete for RP router with other candidate RP. The "**no ip pim rp-candidate**" command cancels the configuration of RP.

3) Configure static RP

Command	Explanation
Global Configuration Mode	
ip pim rp-address <*A.B.C.D*> [<*A.B.C.D/M*>] **no ip pim rp-address <*A.B.C.D*> {<*all*>|<*A.B.C.D/M*>}**	The command is the multicast group configuration static RP of the globally or multicast address range. The no form of this command will remove the configuration for the static RP.

4) Configure the cache time of kernel multicast route

Command	Explanation
Global Configuration Mode	
ip multicast unresolved-cache aging-time <value> **no ip multicast unresolved-cache aging-time**	Configure the cache time of kernel multicast route, the no command restores the default value.

4. Disable PIM-SM Protocol

Command	Explanation
Interface Configuration Mode	
no ip pim sparse-mode | no ip pim multicast-routing(Global configuration mode)	To disable the PIM-SM protocol.

5.3.3 PIM-SM Configuration Examples

As shown in the following figure, add the Ethernet interfaces of SwitchA, SwitchB, SwitchC and SwitchD to corresponding VLAN, and enable PIM-SM Protocol on each VLAN interface.

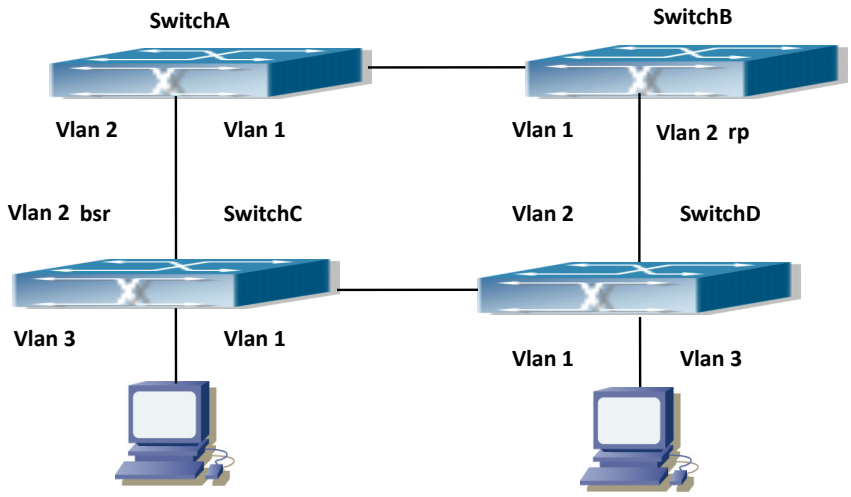


Figure 5-2 PIM-SM Typical Environment

The configuration procedure for SwitchA, SwitchB, SwitchC and SwitchD is as follows:

(1) Configure SwitchA:

Switch(config)#ip pim multicast-routing

Switch(config)#interface vlan 1

Switch(Config-if-Vlan1)# ip address 12.1.1.1 255.255.255.0

Switch(Config-if-Vlan1)# ip pim sparse-mode

Switch(Config-if-Vlan1)#exit

Switch(config)#interface vlan 2

Switch(Config-if-Vlan2)# ip address 13.1.1.1 255.255.255.0

Switch(Config-if-Vlan2)# ip pim sparse-mode

(2) Configure SwitchB:

Switch(config)#ip pim multicast-routing

Switch(config)#interface vlan 1

Switch(Config-if-Vlan1)# ip address 12.1.1.2 255.255.255.0

Switch(Config-if-Vlan1)# ip pim sparse-mode

Switch(Config-if-Vlan1)#exit

Switch(config)#interface vlan 2

Switch(Config-if-Vlan2)# ip address 24.1.1.2 255.255.255.0

Switch(Config-if-Vlan2)# ip pim sparse-mode

Switch(Config-if-Vlan2)# exit

Switch(config)# ip pim rp-candidate vlan2

(3) Configure SwitchC:

Switch(config)#ip pim multicast-routing

Switch(config)#interface vlan 1

Switch(Config-if-Vlan1)# ip address 34.1.1.3 255.255.255.0

Switch(Config-if-Vlan1)# ip pim sparse-mode

Switch(Config-if-Vlan1)#exit

Switch(config)#interface vlan 2

Switch(Config-if-Vlan2)# ip address 13.1.1.3 255.255.255.0

```
Switch(Config-if-Vlan2)# ip pim sparse-mode
Switch(Config-if-Vlan2)#exit
Switch(config)#interface vlan 3
Switch(Config-if-Vlan3)# ip address 30.1.1.1 255.255.255.0
Switch(Config-if-Vlan3)# ip pim sparse-mode
Switch(Config-if-Vlan3)# exit
Switch(config)# ip pim bsr-candidate vlan2 30 10
```

(4) Configure SwitchD:

```
Switch(config)#ip pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)# ip address 34.1.1.4 255.255.255.0
Switch(Config-if-Vlan1)# ip pim sparse-mode
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)# ip address 24.1.1.4 255.255.255.0
Switch(Config-if-Vlan2)# ip pim sparse-mode
Switch(Config-if-Vlan2)#exit
Switch(config)#interface vlan 3
Switch(Config-if-Vlan3)# ip address 40.1.1.1 255.255.255.0
Switch(Config-if-Vlan3)# ip pim sparse-mode
```

At the same time, you should pay attention to the configuration of Unicast Routing Protocol, assure that each device can communicate with each other in the network layer, and be able to implement dynamic routing update in virtue of Unicast Routing Protocol.

5.3.4 PIM-SM Troubleshooting

In configuring and using PIM-SM Protocol, PIM-SM Protocol might not operate normally caused by physical connection or incorrect configuration. Therefore, the user should pay attention to the following issues:

- ☞ Assure that physical connection is correct;
- ☞ Assure the Protocol of Interface and Link is UP (use show interface command);
- ☞ Assure that PIM Protocol is enabled in Global Mode (use ip pim multicast-routing);
- ☞ Assure that PIM-SM is configured on the interface (use ip pim sparse-mode);
- ☞ Multicast Protocol requires RPF Check using unicast routing; therefore the correctness of unicast routing must be assured beforehand;
- ☞ PIM-SM Protocol requires supports by RP and BSR, therefore you should use show ip pim bsr-router first to see if there is BSR information. If not, you need to check if there is unicast routing leading to BSR.
- ☞ Use show ip pim rp-hash command to check if RP information is correct; if there is not RP information, you still need to check unicast routing.

If all attempts including Check are made but the problems on PIM-SM can’t be solved yet, then use debug commands such debug pim/debug pim BSR please, and then copy DEBUG

information in 3 minutes and send to Technology Service Center.

5.4 MSDP

5.4.1 Introduction to MSDP

MSDP – Multicast Source Discovery Protocol, is a protocol that can learn information about multicast source in other PIM-SM domain. The RP on which MSDP is configured will advertise the information about the multicast sources in its domain to all the other MSDP entities through SA messages. Thus, all the information about multicast sources in one PIM-SM domain is spread to another. In MSDP, inter-domain information tree is used other than the shared tree. It is required that the multicast routing protocol used for in-domain routing must be PIM-SM.

☞ The work flow for RP in PIM-SM protocol

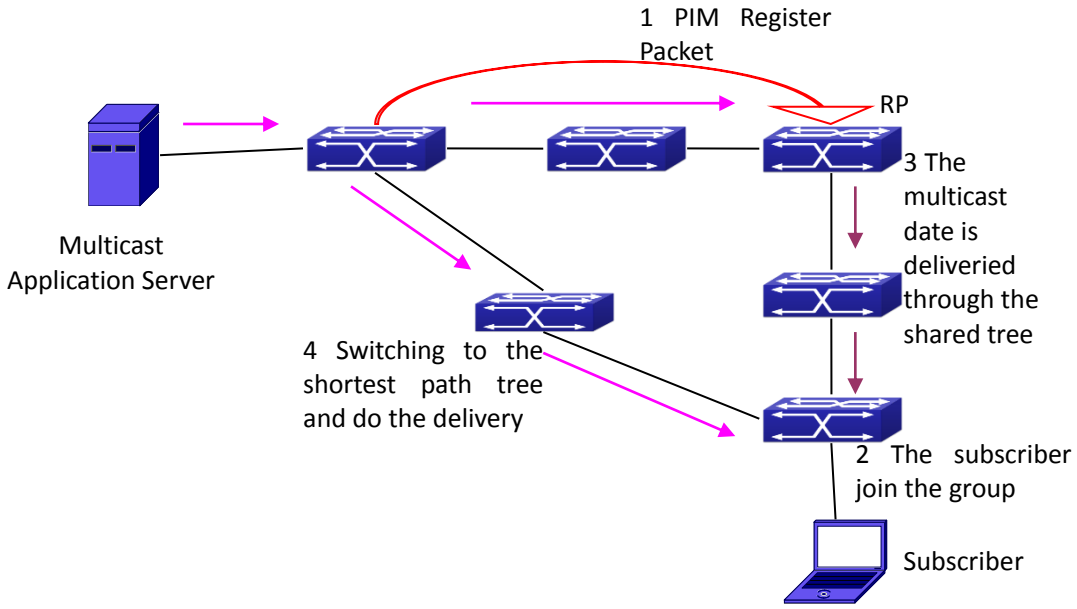


Figure 5-3

5.4.2 Brief Introduction to MSDP Configuration Tasks

1. Configuration of MSDP Basic Function
   1) Enabling MSDP (Required)
   2) Configuring MSDP entities (Required)
   3) Configuring the Connect-Source interface
   4) Configuring static RPF entities
   5) Configuring Originator RP

6) Configuring TTL value

2. Configuration of MSDP entities
   1) Configuring the Connect-Source interface
   2) Configuring the descriptive information for MSDP entities
   3) Configuring the AS number
   4) Configuring the specified mesh group of MSDP
   5) Configuring the maximum size for the cache
3. Configurations on delivery of SA packets
   1) Configuring filter policies for creation of SA packets
   2) Configuring filter rules on how to receive and forward SA packets
   3) Configuring SA request packets
   4) Configuring filter policies for SA-Request packets
4. Configuration of parameters of SA-cache
   1) Configuring SA packets cache
   2) Configuring the aging time for entries in SA packets cache
   3) Configuring the maximum size for the cache

5.4.3 Configuration of MSDP Basic Function

All the commands in this section are configured for RP in the PIM-SM domain. These RP will function as the other peer of the MSDP entities.

5.4.3.1 Prerequisites of MSDP Configuration

Before the MSDP basic functions can be configured, the following tasks should be done:

- At least one single cast routing protocol should be configured, in order to connect the network inside the domain and outside
- Configure PIM-SM in order to implement multicast inside the domain

When configuring MSDP basic function, the following information should be ready:

- The IP address of MSDP entities
- Filter policy table

Pay attention: MSDP can not use with Any-cast RP at same time, but configure Any-cast RP of based MSDP protocol.

5.4.3.2 Enabling MSDP

MSDP should be enabled before various MSDP functions can be configured.

1. Enable the MSDP function
2. Configure MSDP

1. Enabling MSDP

Commands	Explanation
Global Configuration Mode	
router msdp **no router msdp**	To enable MSDP. The no form of this command will disable MSDP globally.

2. Configuration of MSDP parameters

Commands	Explanation
MSDP Configuration Mode	
connect-source ***<interface-type>*** ***<interface-number>*** **no connect-source**	To configure the Connect-Source interface for MSDP Peer. The no form of this command will remove the configured Connect-Source interface.
default-rpf-peer ***<peer-address> [rp-policy <acl-list-number>*****|<word>*** ***]*** ***no*** **default-rpf-peer**	To configure static RPF Peer. The no form of this command will remove the configured RPF Peer.
originating-rp ***<interface-type>*** ***<interface-number>*** **no originating-rp**	To configure Originator-RP. The no form of this command will remove the configured Originator-RP.
ttl-threshold <*ttl*> **no ttl-threshold**	To configure the TTL value. The no form of this command will remove the configured TTL value.

5.4.4 Configuration of MSDP Entities

5.4.4.1 Creation of MSDP Peer

Commands	Explanation
MSDP Configuration Mode	
peer <*peer-address*> **no peer <*peer-address*>**	To create a MSDP Peer. The no form of this command will remove the configured MSDP Peer.

5.4.4.2 Configuration of MSDP parameters

Commands	Explanation
MSDP Peer Configuration Mode	
connect-source ***<interface-type>*** ***<interface-number>*** **no connect-source**	To configure the Connect-Source interface for MSDP Peer. The no form of this command will remove the configured Connect-Source

	interface.
description <*text*> **no description**	To configure the descriptive information about the MSDP entities. The no form of this command will remove the configured description.
remote-as <*as-num*> **no remote-as <*as-num*>**	To configure the AS number for MSDP Peer. The no form of this command will remove the configured AS number of MSDP Peer.
mesh-group <*name*> **no mesh-group <*name*>**	To configure an MSDP Peer to join the specified mesh group. The no form of this command will remove the MSDP Peer from the specified mesh group.

5.4.5 Configuration of Delivery of MSDP Packet

Commands	Explanation
MSDP Configuration Mode	
redistribute [list <*acl-list-number |acl-name*>] **no redistribute**	To configure the filter rules for creation of SA packets. The no form of this command will remove the configured.
MSDP Configuration Mode or MSDP Peer Configuration Mode	
**sa-filter （ in|out ） [list <*acl-number	acl-name*>	rp-list <*rp-acl-number	rp-acl-name*>]** **no sa-filter（in|out）[[list <*acl-number	acl-name*>	rp-list <*rp-acl-number	rp-acl-name*>]**	To configure the filter rules for receiving and forwarding SA packets. The no form of this command will remove the configured rules.
MSDP Peer Configuration Mode	
sa-request **no sa-request**	To configure sending of SA request packets. The no form of this command will disable sending of SA request packets.
MSDP Configuration Mode	
**sa-request-filter [list <*access-list-number	access-list-name*>]** **no sa-request-filter [list <*access-list-number	access-list-name*>]**	To configure filter rules for receiving SA request packets. The no form of this command will remove the configured filter rules for SA request packets.

5.4.6 Configuration of Parameters of SA-cache

Commands	Explanation
MSDP Configuration Mode	

cache-sa-state **no cache-sa-state**	To enable the SA packet cache. To disable the SA packets cache.
MSDP Configuration Mode	
cache-sa-holdtime <*150-3600*> **no cache-sa-holdtime**	The aging time for entries in the SA cache. To restore the default aging time configuration.
MSDP Configuration Mode or MSDP Peer Configuration Mode	
cache-sa-maximum <*sa-limit*> **no cache-sa-maximum**	To configure the maximum size for the SA cache. To restore the size of the SA cache to the default value.

5.4.7 MSDP Configuration Examples

Example 1: MSDP basic function.
Multicast Configuration:

1. Suppose the multicast server is sending multicast datagram at 224.1.1.1;
2. The designated router – DR, which is connected to the multicast server, encapsulate the multicast datagram in the Register packets and send them to the RP(RP1) in the local domain;
3. The RP unwraps the packets and sends them to all the domain members through the shared tree. The members in the domain can be configured to be or not to be in the shared tree;
4. At the same time, the source RP in the domain, generates a SA – Source Active message, and send it to the MSDP entity – RP2.
5. If there's another member in the same domain with the MSDP entity which is named as RP3, RP3 will distribute the multicast datagram encapsulated in the SA messages to the members of the shared tree, and send join messages to the multicast source. That means RP creates an entry (S, G), and send join messages for (S, G) hop by hop, so that (S, G) can reach the SPT which takes the multicast source as the root across the PIM-SM domain.
   If there no members in the same domain with MSDP entity – RP2, RP2 will not create the (S, G) entry nor it will join the SPT which takes the multicast source as the root.
6. When the reverse route has been set up, the multicast datagram from the source will be directly delivered to RP3, and RP will forward the datagram to the shared tree. At this time, the router which is closest to the domain members can determine itself whether or not to switch to SPT.

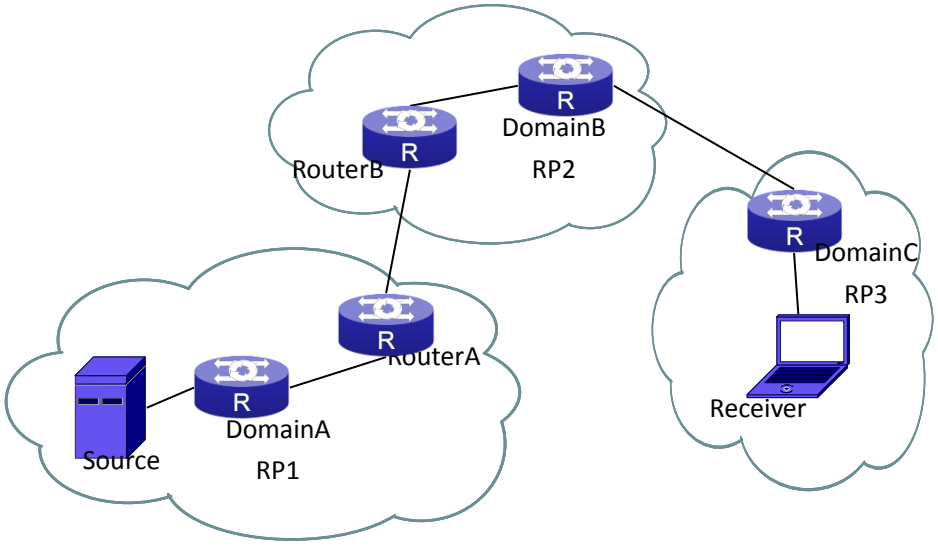


Figure 5-4 Network Topology for MSDP Entry

Configuration tasks are listed as below:

Prerequisites:

Enable the single cast routing protocol and PIM protocol on every router, and make sure that the inter-domain routing works well and multicasting inside the domain works well.

Suppose the multicast server S in Domain A offers multicast programs at 224.1.1.1. A host in Domain C named R subscribes this program. Before MSDP is configured C cannot subscribe the multicast program. However, with the following configuration, R is able to receive programs offered by S.

RP1 in Domain A:

Switch#config
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 10.1.1.1 255.255.255.0
Switch(Config-if-Vlan1)#exit
Switch(config)#router msdp
Switch(router-msdp)#peer 10.1.1.2

Router A in Domain A:

Switch#config
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)#ip address 20.1.1.2 255.255.255.0
Switch(Config-if-Vlan2)#exit
Switch(config)#router msdp
Switch(router-msdp)#peer 10.1.1.1
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 20.1.1.1

Router B in Domain B:

```
Switch#config
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)#ip address 20.1.1.1    255.255.255.0
Switch(Config-if-Vlan2)#exit
Switch(Config)#interface vlan 3
Switch(Config-if-Vlan3)#ip address 30.1.1.1    255.255.255.0
Switch(Config-if-Vlan3)#exit
Switch(config)#router msdp
Switch(router-msdp)#peer 20.1.1.2
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 30.1.1.2
```

RP2 in Domain B:

```
Switch#config
Switch(config)#interface vlan 3
Switch(Config-if-Vlan3)#ip address 30.1.1.2    255.255.255.0
Switch(config)#interface vlan 4
Switch(Config-if-Vlan4)#ip address 40.1.1.2    255.255.255.0
Switch(Config-if-Vlan4)#exit
Switch(config)#router msdp
Switch(router-msdp)#peer 30.1.1.1
Switch(config)#router msdp
Switch(router-msdp)#peer 40.1.1.1
```

RP3 in Domain C:

```
Switch(config)#interface vlan 4
Switch(Config-if-Vlan1)#ip address 40.1.1.1    255.255.255.0
Switch(Config-if-Vlan1)#exit
Switch(config)#router msdp
Switch(router-msdp)#peer 40.1.1.2
```

Example 2: Application of MSDP Mesh-Group.

Mesh-Group can be used to reduce flooding of SA messages. The Peers which are meshed in the same domain can be configured as a Mesh-Group. All the members in the same mesh group use a unique group name.

As it is shown in Figure, when Mesh-Group is configured for the four meshed Peers in the same domain, flooding of SA messages reduced remarkably.

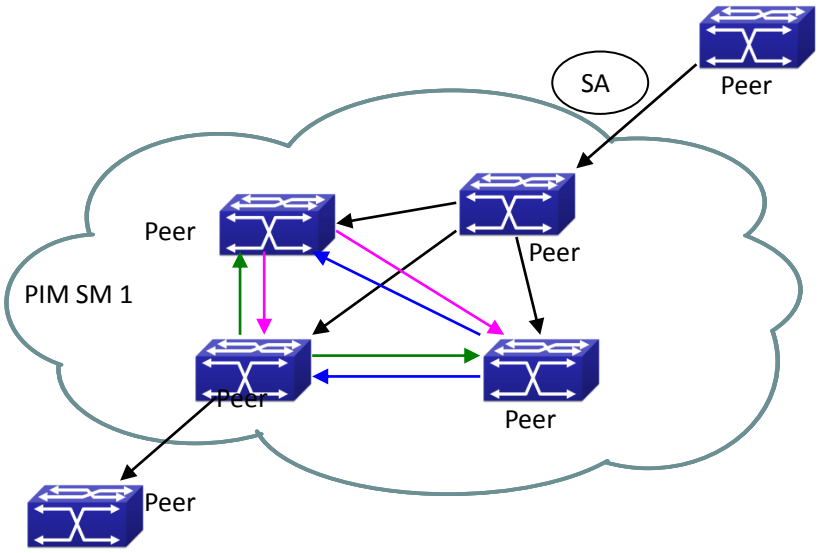


Figure 5-5 Flooding of SA messages

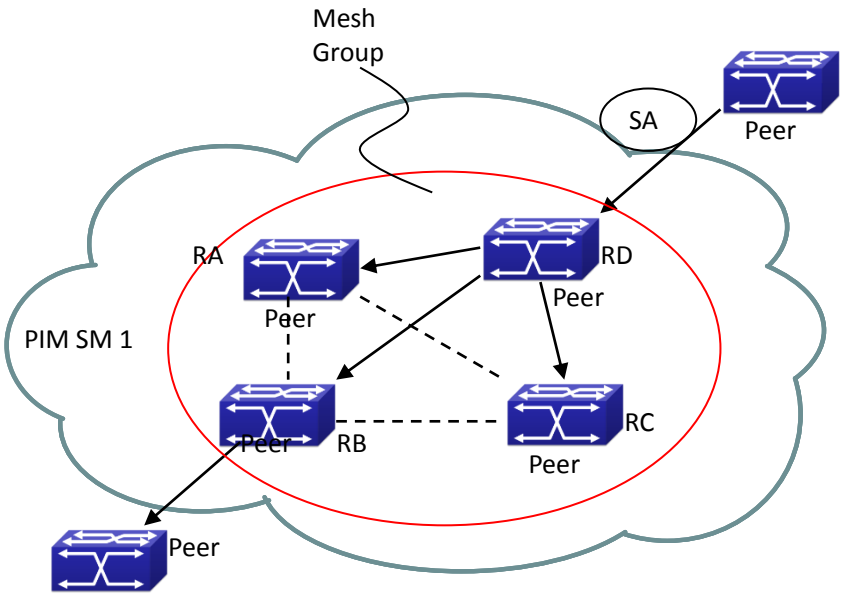


Figure 5-6 Flooding of SA messages with mesh group configuration

Configuration steps are listed as below:

Router A:

Switch#config

Switch(config)#interface vlan 1

Switch(Config-if-Vlan1)#ip address 10.1.1.1 255.255.255.0

Switch(Config-if-Vlan1)#exit

Switch(config)#interface vlan 2

Switch(Config-if-Vlan2)#ip address 20.1.1.1 255.255.255.0

Switch(Config-if-Vlan2)#exit

Switch(config)#interface vlan 3

```
Switch(Config-if-Vlan3)#ip address 30.1.1.1 255.255.255.0
Switch(Config-if-Vlan3)#exit
Switch(config)#router msdp
Switch(router-msdp)#peer 10.1.1.2
Switch(router-msdp)#mesh-group test-1
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 20.1.1.4
Switch(router-msdp)#mesh-group test-1
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 30.1.1.3
Switch(router-msdp)#mesh-group test-1
Switch(msdp-peer)#exit
```

Router B:

```
Switch#config
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 4
Switch(Config-if-Vlan4)#ip address 40.1.1.2 255.255.255.0
Switch(Config-if-Vlan4)#exit
Switch(config)#interface vlan 6
Switch(Config-if-Vlan6)#ip address 60.1.1.2 255.255.255.0
Switch(Config-if-Vlan6)#exit
Switch(config)#router msdp
Switch(router-msdp)#peer 10.1.1.1
Switch(router-msdp)#mesh-group test-1
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 40.1.1.4
Switch(router-msdp)#mesh-group test-1
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 60.1.1.3
Switch(router-msdp)#mesh-group test-1
```

Router C:

```
Switch#config
Switch(config)#interface vlan 4
Switch(Config-if-Vlan4)#ip address 40.1.1.4 255.255.255.0
Switch(Config-if-Vlan4)#exit
Switch(config)#interface vlan 5
Switch(Config-if-Vlan5)#ip address 50.1.1.4 255.255.255.0
Switch(Config-if-Vlan5)#exit
Switch(config)#interface vlan 6
Switch(Config-if-Vlan6)#ip address 60.1.1.4 255.255.255.0
```

```
Switch(Config-if-Vlan6)#exit
Switch(config)#router msdp
Switch(router-msdp)#peer 20.1.1.1
Switch(router-msdp)#mesh-group test-1
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 40.1.1.4
Switch(router-msdp)#mesh-group test-1
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 60.1.1.2
Switch(router-msdp)#mesh-group test-1
```

Router D:

```
Switch#config
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)#ip address 20.1.1.4 255.255.255.0
Switch(Config-if-Vlan2)#exit
Switch(config)#interface vlan 4
Switch(Config-if-Vlan1)#ip address 40.1.1.4    255.255.255.0
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 5
Switch(Config-if-Vlan5)#ip address 50.1.1.4    255.255.255.0
Switch(Config-if-Vlan5)#exit
Switch(config)#router msdp
Switch(router-msdp)#peer 20.1.1.1
Switch(router-msdp)#mesh-group test-1
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 40.1.1.2
Switch(router-msdp)#mesh-group test-1
Switch(msdp-peer)#exit
Switch(router-msdp)#peer 50.1.1.3
Switch(router-msdp)#mesh-group test-1
```

5.4.8 MSDP Troubleshooting

When MSDP is being configured, it may not function because of the physical link not working or configuration mistakes. Attention should be paid to the following items in order to make MSDP work:

- Make sure the physical link works well
- Make sure inner-domain and inter-domain routing works
- Make sure PIM-SM is applied in every domain as the inner-domain routing protocol, and configuration for PIM-SM works well
- Make sure MSDP is enabled, and the link status of the MSDP enabled Peer is UP
- Use the command **show msdp global** to check whether the MSDP configuration is correct

If the MSDP problems cannot be solved through all the methods provided above, please issue the command **debug msdp** to get the debugging messages within three minutes, and send them to the technical service center of our company.

5.5 ANYCAST RP

5.5.1 Introduction to ANYCAST RP

Anycast RP is a technology based on PIM protocol, which provides redundancy in order to recover as soon as possible once an RP becomes unusable.

The kernel concept of Anycast RP is that the RP addresses configured all over the whole network exist on multiple multicast servers (the most common situation is that every device providing ANYCAST RP uses LOOPBACK interface, and using the longest mask to configures RP addresses on this interface), while the unicast routing algorithm will make sure that PIM routers can always find the nearest RP, thus , providing a shorter and faster way to find RP in a larger network., Once an RP being used becomes unusable, the unicast routing algorithm will ensure that the PIM router can find a new RP path fast enough to recover the multicast server in time. Multiple RP will cause a new problem that is if the multicast source and the receivers are registered to different RP, some receivers will not be able to receive data of multicast source (obviously, the register messages only prefer the nearest RP). So, in order to keep the communication between all RP, Anycast RP defines that the nearest RP to the multicast source should forward the source register messages to all the other RP to guarantee that all joiners of the RP can find the multicast source.

The method to realize the PIM-protocol-based Anycast RP is that: maintaining an ANYCAST RP list on every switch configured with Anycast RP and using another address as the label to identify each other. When one Anycast RP device receives a register message, it will send the register message to other Anycast RP devices while using its own address as the source address, to notify all the other devices of the original destination.

5.5.2 ANYCAST RP Configuration Task

1. Enable ANYCAST RP v4 function
2. Configure ANYCAST RP v4

1. Enable ANYCAST RP v4 function

Command	Explanation
Global Configuration Mode	
ip pim anycast-rp **no ip pim anycast-rp**	Enable ANYCAST RP function. (necessary) No operation will globally disable ANYCAST RP function.

2. Configure ANYCAST RP v4

(1) Configure the RP candidate

Command	Explanation
Global Configuration Mode	
ip pim rp-candidate {vlan<*vlan-id*> |loopback<*index*> |<*ifname*>} [<*A.B.C.D*>] [<*priority*>] **no ip pim rp-candidate**	Now, the PIM-SM has allowed the Loopback interface to be a RP candidate.(necessary) Please pay attention to that, ANYCAST RP protocol can configure the Loopback interface or a regular three-layer VLAN interface to be the RP candidate. In make sure that PIM routers in the network can find where the RP locates, the RP candidate interface should be added into the router. No operation will cancel the RP candidate configuration on this router.

(2) Configure self-rp-address (the RP address of this router)

Command	Explanation
Global Configuration Mode	
ip pim anycast-rp self-rp-address A.B.C.D **no ip pim anycast-rp self-rp-address**	Configure the self-rp-address of this router (as a RP). This address can be used to exclusively identify this router when communicating with other RP. the effect of **self-rp-address** refers to two respects: 1 Once this router (as a RP) receives the register message from DR unicast, it needs to forward the register message to all the other RP in the network, notifying them of the state of source (S.G). While forwarding the register message, this router will change the source address of it into self-rp-address. 2 Once this router(as a RP) receives a register message from other RP unicast, such as a register message whose destination is the self-rp-address of this router, it will create (S,G) state and send back a register-stop message, whose destination address is the source address of the register message. Pay attention: self-rp-address has to be the address of a three-layer interface on this router, but the configuration is allowed to be done with the absence of the interface. The self-rp-address should be unique.

	No operation will cancel the self-rp-address which is used to communicate with other RPs by this router (as a RP).

(3) Configure other-rp-address (other RP communication addresses)

Command	Explanation
Global Configuration Mode	
ip pim anycast-rp <*anycast-rp-addr*> <*other-rp-addr*> **no ip pim anycast-rp <*anycast-rp-addr*> <*other-rp-addr*>**	Configure anycast-rp-addr on this router (as a RP). This unicast address is actually the RP address configured on multiple RP in the network, in accordance with the address of RP candidate interface (or Loopback interface). The effect of **anycast-rp-addr** includes: 1 Although more than one anycast-rp-addr addresses are allowed to be configured, only the one having the same address with the currently configured RP candidate address will take effect. Only after that, can the other-rp-address in accordance with this anycast-rp-addr take effect. 2 The configuration is allowed to be done with the absence of the interface in accordance with the anycast-rp-addr. Configure on this router (as a RP) the other-rp-addresses of other RP communicating with it. This unicast address identifies other RP and is used in the communication with local routers. The effect of **other-rp-address** refers to two respects: 1 Once this router (as a RP) receives the register message from a DR unicast, it should forward it to other RP in the network to notify all the RP in the network of the source (S.G) state. While forwarding, the router will change the destination address of the register message into other-rp-address. 2 Multiple other-rp-addresses can be configured in accordance with one anycast-rp-addr, Once the register message from a DR is received, it should be forwarded to all of these other RP one by one. No operation will cancel an other-rp-address communicating with this router.

5.5.3 ANYCAST RP Configuration Examples

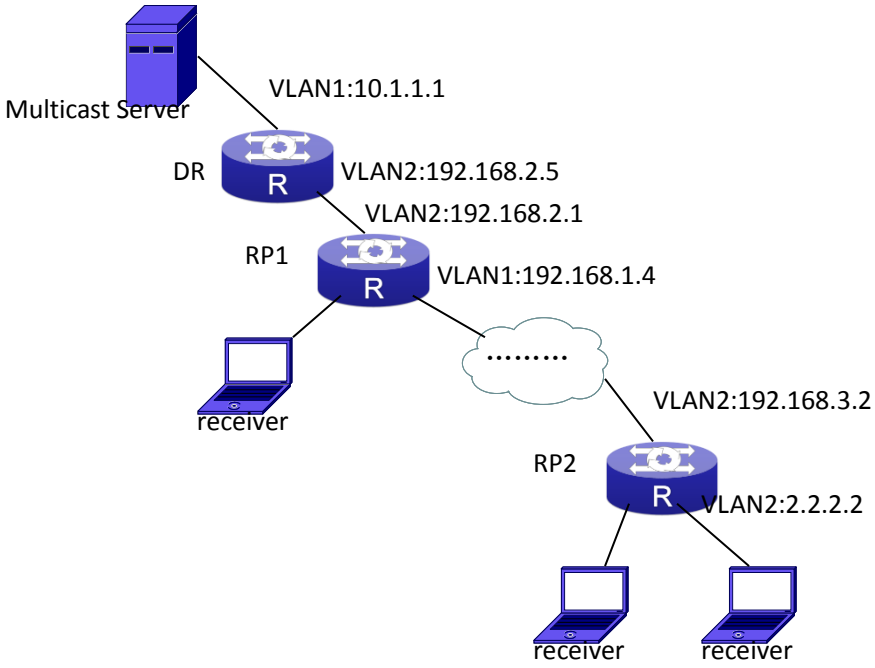


Figure 5-7 The ANYCAST RP v4 function of the router

As shown in the Figure, the overall network environment is PIM-SM, which provides two routers supporting ANYCAST RP, RP1 and RP2. Once multicast data from the multicast source server reaches the DR, the DR will send a multicast source register message to the nearest RP unicast according to the unicast routing algorithm, which is RP1 in this example. When RP1 receives the register message from the DR, besides redistributing to the shared tree according to the orderers who already join it, it will forward the multicast register message to RP2 to guarantee that all orders that already join RP2 can find the multicast source. Since there is an ANYCAST list maintained on router RP1 that has been configured with ANYCAST RP, and since this list contains the unicast addresses of all the other RP in the network, when the RP1 receives the register message, it can use the self-r-address, which identifies itself as the source address to forward the register message to RP2. The cloud in the Figure represents the PIM-SM network operation between RP1 and RP2.

The following is the configuration steps:

RP1 Configuration:

Switch#config
Switch(config)#interface loopback 1
Switch(Config-if-Loopback1)#ip address 1.1.1.1 255.255.255.255
Switch(Config-if-Loopback1)#exit
Switch(config)#ip pim rp-candidate loopback1
Switch(config)#ip pim bsr-candidate vlan 1
Switch(config)#ip pim multicast-routing
Switch(config)#ip pim anycast-rp
Switch(config)#ip pim anycast-rp self-rp-address 192.168.2.1

Switch(config)#ip pim anycast-rp 1.1.1.1 192.168.3.2

RP2 Configuration:

Switch#config

Switch(config)#interface loopback 1

Switch(Config-if-Loopback1)#ip address 1.1.1.1 255.255.255.255

Switch(Config-if-Loopback1)#exit

Switch(config)#ip pim rp-candidate loopback1

Switch(config)#ip pim multicast-routing

Switch(config)#ip pim anycast-rp

Switch(config)#ip pim anycast-rp self-rp-address 192.168.3.2

Switch(config)#ip pim anycast-rp 1.1.1.1 192.168.2.1

5.5.4 ANYCAST RP Troubleshooting

When configuring and using ANYCAST RP function, the ANYCAST RP might work abnormally because of faults in physical connections, configurations or something others. So, the users should pay attention to the following points:

- The physical connections should be guaranteed to be correct
- The PIM-SM protocol should be guaranteed to operate normally
- The ANYCAST RP should be guaranteed to be enabled in Global configuration mode
- The self-rp-address should be guaranteed to be configured correctly in Global configuration mode
- The other-rp-address should be guaranteed to be configured correctly in Global configuration mode
- All the interface routers should be guaranteed to be correctly added, including the loopback interface as a RP
- Use "**show ip pim anycast rp status**" command to check whether the configuration information of ANYCAST RP is correct

If the problems of ANYCAST still cannot be solved after checking, please use debug commands like "**debug pim anycast-rp**", then copy the DEBUG information within three minutes and send it to the technical service center of our company.

5.6 PIM-SSM

5.6.1 Introduction to PIM-SSM

Source Specific Multicast (PIM-SSM) is a new kind of multicast service protocol. With PIM-SSM, a multicast session is distinguished by the multicast group address and multicast source address. In SSM, hosts can be added into the multicast group manually and efficiently like

the traditional PIM-SM, but leave out the shared tree and RP management in PIM-SM. In SSM, SPT tree will be constructed with (S, G). G for the multicast group address and S for the source address of the multicast which sends datagram to G. (S, G) in a pair is named as a channel of SSM. SSM serves best for the application of multicast service which is from one station to many ones, for example, the network sports video channel, and the news channel. By default, the multicast group address of SSM is limited between 232.0.0.0 and 232.255.255.255. However this address range can be extended according to actual situations.

5.6.2 PIM-SSM Configuration Task List

Command	Explanation
Global Configuration Mode	
ip multicast ssm {default|range <*access-list-number*>} **no ip multicast ssm**	To configure the address range for pim-ssm. The no form command will disable the configuration.

5.6.3 PIM-SSM Configuration Examples

As the figure shows, ethernet interfaces from SwitchA, SwitchB, SwitchC, and SwitchD are configured to be in separate VLANs. And PIM-SSM is enabled globally by enabling the PIM-SM or PIM-DM protocol on the VLAN interfaces. Take PIM-SM for example.

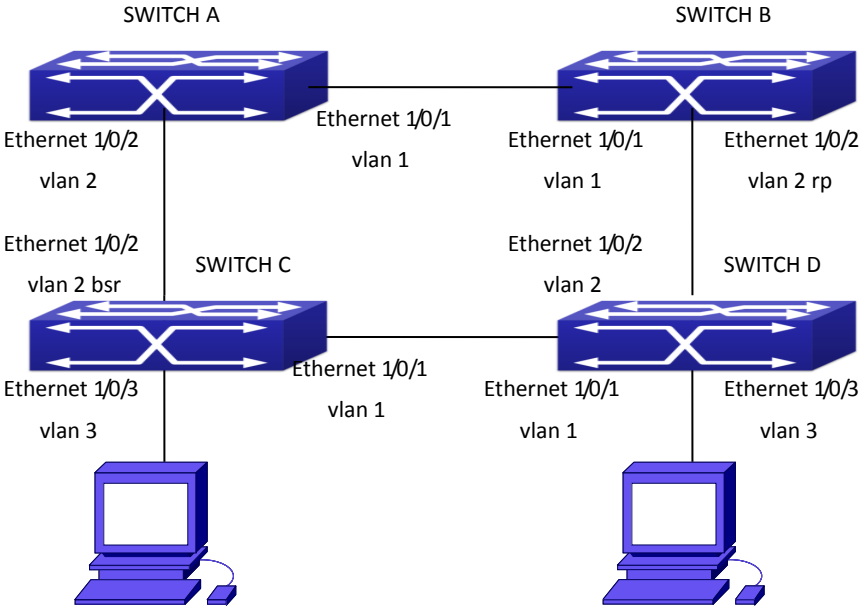


Figure 5-8 PIM-SSM typical environment

Configurations of SwitchA, SwitchB, SwitchC, and SwitchD are shown as below.

(1) Configuration of SwitchA.

Switch(config)#ip pim multicast-routing

```
Switch(config)#interface vlan 1
Switch(Config-If-Vlan1)# ip pim sparse-mode
Switch(Config-If-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-If-Vlan2)# ip pim sparse-mode
Switch(Config-If-Vlan2)#exit
Switch(config)#access-list 1 permit 224.1.1.1 0.0.0.255
Switch(config)#ip multicast ssm range 1
```

(2) Configuration of SwitchB.

```
Switch(config)#ip pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-If-Vlan1)# ip pim sparse-mode
Switch(Config-If-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-If-Vlan2)# ip pim sparse-mode
Switch(Config-If-Vlan2)# exit
Switch(config)# ip pim rp-candidate vlan2
Switch(config)#access-list 1 permit 224.1.1.1 0.0.0.255
Switch(config)#ip multicast ssm range 1
```

(3) Configuration of SwitchC.

```
Switch(config)#ip pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-If-Vlan1)# ip pim sparse-mode
Switch(Config-If-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-If-Vlan2)# ip pim sparse-mode
Switch(Config-If-Vlan2)#exit
Switch(config)#interface vlan 3
Switch(Config-If-Vlan3)# ip pim sparse-mode
Switch(Config-If-Vlan3)# exit
Switch(config)# ip pim bsr-candidate vlan2 30 10
Switch(config)#access-list 1 permit 224.1.1.1 0.0.0.255
Switch(config)#ip multicast ssm range 1
```

(4) Configuration of SwitchD.

```
Switch(config)#ip pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-If-Vlan1)# ip pim sparse-mode
Switch(Config-If-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-If-Vlan2)# ip pim sparse-mode
Switch(Config-If-Vlan2)#exit
Switch(config)#interface vlan 3
Switch(Config-If-Vlan3)# ip pim sparse-mode
Switch(Config-If-Vlan3)#exit
```

```
Switch(config)#access-list 1 permit 224.1.1.1 0.0.0.255
Switch(config)#ip multicast ssm range 1
```

5.6.4 PIM-SSM Troubleshooting

In configuring and using PIM-SSM Protocol, PIM-SSM Protocol might not operate normally caused by physical connection or incorrect configuration. Therefore, the user should pay attention to the following issues:

- Assure that physical connection is correct;
- Assure the Protocol of Interface and Link is UP (use **show interface** command);
- Assure that PIM Protocol is enabled in Global Mode (use **ip pim multicast-routing**);
- Assure that PIM-SSM is configured on the interface (use **ip pim sparse-mode**);
- Assure that SSM is configured in Global Mode;
- Multicast Protocol requires RPF check using unicast routing, therefore the correctness of unicast routing must be assured beforehand.

If all attempts including check are made but the problems on PIM-SSM can't be solved yet, then use debug commands such **debug pim event/debug pim packet** please, and then copy DEBUG information in 3 minutes and send to Technology Service Center.

5.7 DVMRP

5.7.1 Introduction to DVMRP

DVMRP Protocol, namely, is "Distance Vector Multicast Routing Protocol". It is a Multicast Routing Protocol in dense mode, which sets up a Forward Broadcast Tree for each source in a manner similar to RIP, and sets up a Truncation Broadcast Tree, i.e. the Shortest Path Tree to the source, for each source through dynamic Prune/Graft.

Some of the important features of DVMRP are:

1. The routing exchange used to determine reverse path checking information is based on distance vector (in a manner similar to RIP)
2. Routing exchange update occurs periodically (the default is 60 seconds)
3. TTL upper limit = 32 hops (and that RIP is 16)
4. Routing update includes net mask and supports CIDR

In comparison with Unicast routing, Multicast routing is a kind of reverse routing (that is, what you are interested in is where the packets are from but not where they go), thus the information in DVMRP routing table is used to determine if an input Multicast packet is received at the correct interface. Otherwise, the packet will be discarded to prevent Multicast circulation.

The check which determines if the packet gets to the correct interface is called RPF check. When some Multicast data packets get to some interface, it will determine the reverse path to

the source network by looking up DVMRP router table. If the interface data packets get to is the one which is used to send Unicast message to the source, then the reverse path check is correct, and the data packets are forwarded out from all downstream interfaces. If not, then probably there is failure, and the Multicast packet is discarded.

Since not all switches support Multicast, DVMRP supports tunnel multicast communication, tunnel is a method to send multicast data report among DVMRP switches separated by switches which don't support multicast routing. Multicast data packets are encapsulated in unicast data packets and directly sent to the next switch which supports multicast. DVMRP Protocol treats tunnel interface and general physical interface equally.

If two or more switches are connected to a multi-entrance network, it is likely to transmit more than one copy of a data packet to the sub-network. Thus a specified transmitter must be appointed. DVMRP achieves this goal by making use of routing exchange mechanism; when two switches on the multi-entrance network exchange routing information, they will be aware of the routing distance from each other to the source network, thus the switch with the shortest distance to the source network will become the specified transmitter of the sub-network. If some have the same distance, then the one with the lowest IP prevails.

After some interface of the switch is configured to Function DVMRP Protocol, the switch will multicast Probe message to other DVMRP switches on this interface, which is used to find neighbors and detect the capabilities of each other. If no Probe message from the neighbor is received until the neighbor is timed out, then this neighbor is considered missing.

In DVMRP, source network routing selection message are exchanged in a basic manner same to RIP. That is, routing report message is transmitted among DVMRP neighbors periodically (the default is 60 seconds). The routing information in DVMRP routing selection table is used to set up source distribution tree, i.e. to determine by which neighbor it passes to get to the source transmitting multicast packet; the interface to this neighbor is called upstream interface. The routing report includes source network (use net mask) address and the hop entry for routing scale.

In order to finish transmission correctly, every DVMRP switch needs to know which downstream switches need to receive multicast packet from some specific source network through it. After receiving packets from some specific source, DVMRP switch firstly will broadcast these multicast packets from all downstream interfaces, i.e. the interfaces on which there are other DVMRP switches which have dependence on the specific source. After receiving Prune message from some downstream switch on the interface, it will prune this switch. DVMRP switch uses poison reverse to notify the upstream switch for some specific source: "I am your downstream." By adding infinity (32) to the routing distance of some specific source it broadcasts, DVMRP switch responds to the source upstream exchange to fulfill poison reverse. This means distance correct value is 1 to 2* infinity (32) -1 or 1 to 63, 1 to 63 means it can get to source network, 32 means source network is not arrival, 33 to 63 means the switch which generates the report message will receive multicast packets from specific source depending on upstream router.

5.7.2 DVMRP Configuration Task List

1. Globally enable and disable DVMRP (Required)
2. Configure Enable and Disable DVMRP Protocol at the interface (Required)

3. Configure DVMRP Sub-parameters (Optional)
   Configure DVMRP interface parameters
   1) Configure the delay of transmitting report message on DVMRP interface and the message number each time it transmits
   2) Configure metric value of DVMRP interface
   3) Configure if DVMRP is able to set up neighbors with DVMRP routers which can not Prune/Graft
4. Configure DVMRP tunnel

1. Globally enable DVMRP Protocol

The basic configuration to function DVMRP routing protocol on Layer 3 switch is very simple. Firstly it is required to turn on DVMRP switch globally.

Command	Explanation
Global Mode	
[no] ip dvmrp multicast-routing	Globally enable DVMRP Protocol, the "**no ip dvmrp multicast-routing**" command disables DVMRP Protocol globally. (Required)

2. Enable DVMRP Protocol on the interface

The basic configuration to function DVMRP routing protocol on Layer 3 switch is very simple. After globally enabling DVMRP Protocol, it is required to turn on DVMRP switch under corresponding interface.

Command	Explanation
Interface Configuration Mode	
ip dvmrp **no ip dvmrp**	Enable DVMRP Protocol on the interface, the "**no ip dvmrp**" command disables DVMRP Protocol on the interface.

3. Configure DVMRP Sub-parameters

（1）Configure DVMRP Interface Parameters

   1) Configure the delay of transmitting report message on DVMRP interface and the message number each time it transmits
   2) Configure metric value of DVMRP interface
   3) Configure if DVMRP is able to set up neighbors with DVMRP routers which can not Prune/Graft

Command	Explanation
Interface Configuration Mode	
ip dvmrp output-report-delay <*delay_val*> [<*burst_size*>] **no ip dvmrp output-report-delay**	Configure the delay of transmitting DVMRP report message on interface and the message number each time it transmits, the "**no ip dvmrp output-report-delay**" command restores default value.

ip dvmrp metric <*metric_val*> **no ip dvmrp metric**	Configure interface DVMRP report message metric value; the "**no ip dvmrp metric**" command restores default value.
ip dvmrp reject-non-pruners **no ip dvmrp reject-non-pruners**	Configure the interface rejects to set up neighbor relationship with non pruning/grafting DVMRP router. The "**no ip dvmrp reject-non-pruners**" command restores to being able to set up neighbor ship.

4. Configure DVMRP Tunnel

Command	Explanation
Interface Configuration Mode	
ip dvmrp tunnel *<index>* *<src-ip>* *<dst-ip>* **no ip dvmrp tunnel {*<index>* |*<src-ip>* *<dst-ip>*}**	This command configures a DVMRP tunnel; the "**no ip dvmrp tunnel {*<index>* |*<src-ip>* *<dst-ip>*}**" command deletes a DVMRP tunnel.

5.7.3 DVMRP Configuration Examples

As shown in the following figure, add the Ethernet interfaces of Switch A and Switch B to corresponding VLAN, and enable DVMRP on each VLAN interface.

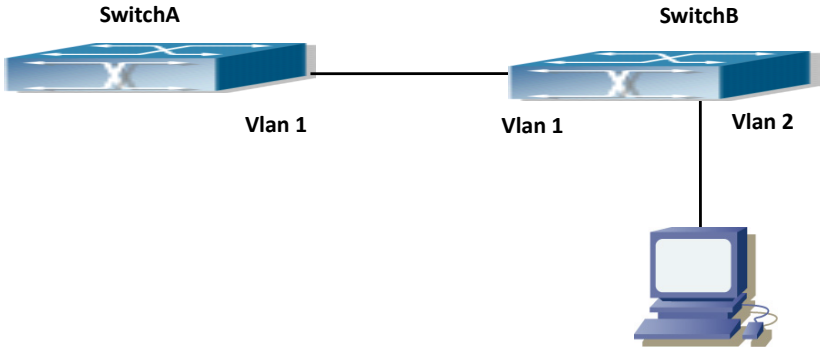


Figure 5-9 DVMRP Network Topology Diagram

The configuration procedure for SwitchA and SwitchB is as follows:

(1) Configure SwitchA:

Switch (config)#ip dvmrp multicast-routing

Switch (config)#interface vlan 1

Switch(Config-if-Vlan1)# ip address 10.1.1.1 255.255.255.0

Switch(Config-if-Vlan1)# ip dvmrp enable

(2) Configure SwitchB:

Switch (config)#ip dvmrp multicast-routing

Switch (config)#interface vlan 1

Switch(Config-if-Vlan1)# ip address 12.1.1.2 255.255.255.0

Switch(Config-if-Vlan1)# ip dvmrp enable

Switch(Config-if-Vlan1)#exit

```
Switch (config)#interface vlan 2
Switch(Config-if-Vlan2)# ip address 20.1.1.1 255.255.255.0
Switch(Config-if-Vlan2)# ip dvmrp
```

Since DVMRP itself does not rely on Unicast Routing Protocol, it is not necessary to configure Unicast Routing Protocol. This is the difference from PIM-DM and PIM-SM.

5.7.4 DVMRP Troubleshooting

In configuring and using DVMRP Protocol, DVMRP Protocol might not operate normally caused by physical connection or incorrect configuration. Therefore, the user should pay attention to the following issues:

- Firstly to assure that physical connection is correct;
- Next, to assure the Protocol of Interface and Link is UP (use **show interface** command);
- Please check if the correct IP address is configured on the interface (use **ip address** command);
- Afterwards, enable DVMRP Protocol on the interface (use **ip dvmrp** command and **ip dv multicast-routing** command);
- Multicast Protocol requires RPF Check using unicast routing; therefore the correctness of unicast routing must be assured beforehand. (DVMRP uses its own unicast table, please use **show ip dvmrp route** command to look up).

If all attempts including Check are made but the problems on DVMRP can't be solved yet, then please use commands such as debug DVMRP, and then copy DEBUG information in 3 minutes and send to Technology Service Center.

5.8 DCSCM

5.8.1 Introduction to DCSCM

DCSCM (Destination control and source control multicast) technology mainly includes three aspects, i.e. Multicast Packet Source Controllable, Multicast User Controllable and Service-Oriented Priority Strategy Multicast.

The Multicast Packet Source Controllable technology of Security Controllable Multicast technology is mainly processed in the following manners:

1. On the edge switch, if source under-control multicast is configured, then only multicast data from specified group of specified source can pass.
2. For RP switch in the core of PIM-SM, for REGISTER information out of specified source and specified group, REGISTER_STOP is transmitted directly and table entry is not allowed to set up. (This task is implemented in PIM-SM model).

The implement of Multicast User Controllable technology of Security Controllable

Multicast technology is based on the control over IGMP report message sent out by the user, thus the model being controlled is IGMP snooping and IGMPmodel, of which the control logic includes the following three, i.e. to take control based on VLAN+MAC address transmitting packets, to take control based on IP address of transmitting packets and to take control based on the port where messages enter, in which IGMP snooping can use the above three methods to take control simultaneously, while since IGMP model is located at layer 3, it only takes control over the IP address transmitting packets.

The Service-Oriented Priority Strategy Multicast of Security Controllable technology adopts the following mode: for multicast data in limit range, set the priority specified by the user at the join-in end so that data can be sent in a higher priority on TRUNK port, consequently guarantee the transmission is processed in user-specified priority in the entire network.

5.8.2 DCSCM Configuration Task List

1. Source Control Configuration
2. Destination Control Configuration
3. Multicast Strategy Configuration

1. Source Control Configuration

Source Control Configuration has three parts, of which the first is to enable source control. The command of source control is as follows:

Command	Explanation
Global Configuration Mode	
[no] ip multicast source-control (Required)	Enable source control globally, the "**no ip multicast source-control**" command disables source control globally. It is noticeable that, after enabling source control globally, all multicast packets are discarded by default. All source control configuration can not be processed until that it is enabled globally, while source control can not be disabled until all configured rules are disabled.

The next is to configure the rule of source control. It is configured in the same manner as for ACL, and uses ACL number of 5000-5099, every rule number can be used to configure 10 rules. It is noticeable that these rules are ordered, the front one is the one which is configured the earliest. Once the configured rules are matched, the following rules won't take effect, so rules of globally allow must be put at the end. The commands are as follows:

Command	Explanation
Global Configuration Mode	

[no] access-list <*5000-5099*> {deny|permit} ip {{<*source*> <*source-wildcard*>}|{host-source <*source-host-ip*>}|any-source} {{<*destination*> <*destination-wildcard*>}|{host-destination <*destination-host-ip*>}|any-destination}	The rule used to configure source control. This rule does not take effect until it is applied to specified port. Using the NO form of it can delete specified rule.

The last is to configure the configured rule to specified port.

Note: If the rules being configured will occupy the table entries of hardware, configuring too many rules will result in configuration failure caused by bottom table entries being full, so we suggest user to use the simplest rules if possible. The configuration rules are as follows:

Command	Explanation
Port Configuration Mode	
[no] ip multicast source-control access-group <*5000-5099*>	Used to configure the rules source control uses to port, the NO form cancels the configuration.

2. Destination Control Configuration

Like source control configuration, destination control configuration also has three steps.

First, enable destination control globally. Since destination control need to prevent unauthorized user from receiving multicast data, the switch won't broadcast the multicast data it received after configuring global destination control. Therefore, It should be avoided to connect two or more other Layer 3 switches in the same VLAN on a switch on which destination control is enabled. The configuration commands are as follows:

Command	Explanation
Global Configuration Mode	
[no] multicast destination-control（required）	Globally enable IPv4 and IPv6 destination control. The no operation of this command will globally disable destination control. All of the other configuration can only take effect after globally enabled. The next is configuring destination control rules, which are similar.

Next is to configure the multicast destination control profile rule list and use the profile-id number of 1-50.

Command	Explanation
Global Configuration Mode	

profile-id <1-50> {deny|permit} {{<source/M> }|{host-source <source-host-ip> (range <2-65535>|)}|any-source} {{<destination/M>}|{host-destination <destination-host-ip> (range <2-255>|)}|any-destination} **no profile-id <1-50>**	Configure the destination control profile rule. The no command deletes it.

Then configure destination control rule. It is similar to source control, except to use ACL No. of 6000-7999.

Command	Explanation
Global Configuration Mode	
[no] access-list <6000-7999> {{{add | delete} profile-id WORD} | {{deny|permit} (ip) {{<source/M> }|{host-source <source-host-ip> (range <2-65535>|)}|any-source} {{<destination/M>}|{host-destination <destination-host-ip> (range <2-255>|)}|any-destination}}	The rule used to configure destination control. This rule does not take effect until it is applied to source IP or VLAN-MAC and port. Using the NO form of it can delete specified rule.

The last is to configure the rule to specified source IP, source VLAN MAC or specified port. It is noticeable that, due to the above situations, these rules can only be used globally in enabling IGMP-SNOOPING. And if IGMP-SNOOPING is not enabled, then only source IP rule can be used under IGMP Protocol. The configuration commands are as follows:

Command	Explanation
Port Configuration Mode	
[no] ip multicast destination-control access-group <*6000-7999*>	Used to configure the rules destination control uses to port, the NO form cancels the configuration.
Global Configuration Mode	
[no] ip multicast destination-control <*1-4094*> <*macaddr*> access-group <6000-7999>	Used to configure the rules destination control uses to specify VLAN-MAC, the NO form cancels the configuration.
[no] ip multicast destination-control <*IPADDRESS/M*> access-group <*6000-7999*>	Used to configure the rules destination control uses to specified IP address/net mask, the NO form cancels the configuration.

3. Multicast Strategy Configuration

Multicast Strategy uses the manner of specifying priority for specified multicast data to achieve and guarantee the effects the specific user requires. It is noticeable that multicast data can not get a special care all along unless the data are transmitted at TRUNK port. The configuration is very simple, it has only one command, i.e. to set priority for the specified multicast. The commands are as follows:

Command	Explanation

Global Configuration Mode	
[no] ip multicast policy <*IPADDRESS/M*> <*IPADDRESS/M*> cos <*priority*>	Configure multicast strategy, specify priority for sources and groups in specific range, and the range is <0-7>.

5.8.3 DCSCM Configuration Examples

1. Source Control

In order to prevent an Edge Switch from putting out multicast data ad asbitsium, we configure Edge Switch so that only the switch at port Ethernet1/0/5 is allowed to transmit multicast, and the data group must be 225.1.2.3. Also, switch connected up to port Ethernet1/0/10 can transmit multicast data without any limit, and we can make the following configuration.

EC(config)#access-list 5000 permit ip any host 225.1.2.3

EC(config)#access-list 5001 permit ip any any

EC(config)#ip multicast source-control

EC(config)#interface ethernet1/0/5

EC(Config-If-Ethernet1/0/5)#ip multicast source-control access-group 5000

EC(config)#interface ethernet1/0/10

EC(Config-If-Ethernet1/0/10)#ip multicast source-control access-group 5001

2. Destination Control

We want to limit users with address in 10.0.0.0/8 network segment from entering the group of 238.0.0.0/8, so we can make the following configuration:

Firstly enable IGMP snooping in the VLAN it is located (Here it is assumed to be in VLAN2)

EC(config)#ip igmp snooping

EC(config)#ip igmp snooping vlan 2

After that, configure relative destination control access-list, and configure specified IP address to use that access-list.

Switch(config)#access-list 6000 deny ip any 238.0.0.0 0.255.255.255

Switch(config)#access-list 6000 permit ip any any

Switch(config)#multicast destination-control

Switch(config)#ip multicast destination-control 10.0.0.0/8 access-group 6000

In this way, users of this network segment can only join groups other than 238.0.0.0/8.

Or configure the destination control access-list by adding the profile list.

Switch (config)#profile-id 1 deny ip any 238.0.0.0 0.255.255.255

Switch (config)#access-list 6000 add profile-id 1

Switch (config)#multicast destination-control

Switch (config)#ip multicast destination-control 10.0.0.0/8 access-group 6000

3. Multicast strategy

Server 210.1.1.1 is distributing important multicast data on group 239.1.2.3, we can configure on its join-in switch as follows:

Switch(config)#ip multicast policy 210.1.1.1/32 239.1.2.3/32 cos 4

In this way, the multicast stream will have a priority of value 4 (Usually this is pretty higher, the higher possible one is protocol data; if higher priority is set, when there is too many multicast

data, it might cause abnormal behavior of the switch protocol) when it gets to other switches through this switch.

5.8.4 DCSCM Troubleshooting

The effect of DCSCM module itself is similar to ACL, and the problems occurred are usually related to improper configuration. Please read the descriptions above carefully. If you still can not determine the cause of the problem, please send your configurations and the effects you expect to the after-sale service staff of our company.

5.9 IGMP

5.9.1 Introduction to IGMP

IGMP (Internet Group Management Protocol) is the protocol in TCP/IP protocol family which is responsible for IP multicast member management. It is used to set up and maintain multicast group member relationship between IP host and its neighbor multicast switches. IGMP does not include the spread and maintenance of relation information of group members among multicast switches, this work is accomplished by each multicast routing protocol. All hosts participating in multicast must implement IGMP protocol.

Hosts participating IP multicast can join in and exit multicast group at any location, any time and without limit of member total. Multicast switch does not need and not likely to save all relationships of all hosts. It only gets to know if there are receivers of some multicast group, i.e. group member, on the network segment each interface connects to. And the host only needs to save which multicast groups it joined.

IGMP is asymmetric between host and router: the host needs to respond the IGMP query messages of multicast switches, i.e. to report message response in membership; the switch sends out membership query messages periodically, and then determine if there are hosts of some specific group joining in the sub-network it belongs to based on the received response message, and send out query of specific group (IGMP version2) when receiving the report of a host exiting the group to determine if there exists no member in some specific group.

Up to now, there are three versions of IGMP: IGMP version1 (defined by RFC1112), IGMP version2 (defined by RFC2236) and IGMP version3 (defined by RFC3376).

The main improvements of IGMP version2 over version1 are:

1. The election mechanism of multicast switches on the shared network segment

Shared network segment is the situation of there is more than one multicast switch on a network segment. Under this kind of situation, since all switches which runs IGMP under this network segment can get membership report message from the host, therefore, only one switch is required to transmit membership query message, so an exchange election mechanism is required to determine a switch as query machine. In IGMP version1, the selection of query machine is determined by Multicast Routing Protocol; IGMP version2 made an improvement for

it, it prescribed that when there are more than one multicast switches on the same network segment, the multicast switch with the lowest IP address will be elected as the query machine.

2. IGMP version2 added Leave Group Mechanism

In IGMP version 1, the host leaves the multicast group silently without sending any notification to any multicast switch. This causes that the multicast switch can only determine the leave of multicast member by multicast group response time-out. But in version2, when a host decides to leave a multicast group, if it is the host which gives response to the latest membership query message, then it will send out a message implying it is leaving.

3. IGMP version 2 added the query to specific group

In IGMP version1, a query of multicast switch is for all multicast groups on the network segment. This query is called general group query. In IGMP version2, query of specific group is added besides general group query. The destination IP address of this kind of query message is the IP address of the multicast group, the group address field part of the message is also the IP address of the multicast group. Thus it is prevented that hosts which are other multicast group members transmit response message.

4. IGMP version2 added the biggest response time field

IGMP version2 added the biggest response time field to dynamically adjust the response time of the host to group query message.

The main features of version3 is allowing the host to choose receiving from or rejecting a certain source, which is the basis of SSM （Source-Specific Multicast）multicast. For example, when a host is sending a report of INCLUDE{10.1.1.1, 10.1.1.2} to some group G, that means the host needs the router to forward the flux from 10.1.1.1 and 10.1.1.2; when a host is sending a report of EXCLUDE{192.168.1.1} to some group G, that means the host needs the flux from all sources of group G except 192.168.1.1. This makes a great difference from the previous IGMP.

The main improvements of IGMP Version3 over IGMP Version1 and Version2 are:

1. The status to be maintained is group and source list, not only the groups in IGMPv2.
2. The interoperations with IGMPv1 and IGMPv2 are defined in IGMPv3 status.
3. IP service interface is modified to allow specific source list thereby.
4. The queried includes his/her Robustness Variable and Query Interval in query group to allow the synchronization with these variables of non-queries.
5. Max Response Time in Query Message has an exponential range, with maximum value from 25.5 secs of v2 to 53 mins, which can be used in links of great capacity.
6. In order to increase strength, the host retransmits State-Change message.
7. Additional data is defined to adapt future extension.
8. Report group is sent to 224.0.0.22 to help with IGMP Snooping of Layer 2 Switch.
9. Report group can include more than one group record, and it allows using small group to report complete current status.
10. The host does not restrain operation any more, which simplifies the implement and allows direct membership trace.
11. In querying messages, the new router side restraint process (S sign) modified the existing strength of IGMPv2.

5.9.2 IGMP Configuration Task List

1. Enable IGMP (Required)
2. Configure IGMP sub-parameters (Optional)
   （1）Configure IGMP group parameters
      1) Configure IGMP group filtering conditions
      2) Configure IGMP to join in group
      3) Configure IGMP to join in static group
   （2）Configure IGMP query parameters
      1) Configure the interval of IGMP sending query message
      2) Configure the maximum response time of IGMP query
      3) Configure time-out of IGMP query
   （3）Configure IGMP version
3. Disable IGMP Protocol

1. Enable IGMP Protocol

There are not specific commands for enabling IGMP Protocol on the Layer 3 switch. Enabling any multicast protocol under corresponding interface will automatically enable IGMP.

Command	Explanation
Global Mode	
ip dvmrp multicast-routing | ip pim multicast-routing	To enable global multicast protocol is the prerequisite to enable IGMP protocol, the "**no ip dvmrp multicast-routing | no ip pim multicast-routing**" commands disable multicast protocol and IGMP protocol. (Required)

Command	Explanation
Interface Configuration Mode	
ip dvmrp enable| ip pim dense-mode | ip pim sparse-mode	Enable IGMP Protocol, the corresponding commands "**no ip dvmrp enable| no ip pim dense-mode | no ip pim sparse-mode**" disable IGMP Protocol. (Required)

2. Configure IGMP Sub-parameters

（1）Configure IGMP group parameters

1) Configure IGMP group filtering conditions
2) Configure IGMP to join in group
3) Configure IGMP to join in static group

Command	Explanation
Interface Configuration Mode	

**ip igmp access-group {<*acl_num	acl_name*>}** **no ip igmp access-group**	Configure the filtering conditions of the interface to IGMP group; the "**no ip igmp access-group**" command cancels the filtering condition.
ip igmp join-group <*A.B.C.D* > **no ip igmp join-group <*A.B.C.D* >**	Configure the interface to join in some IGMP group, the "**no ip igmp join-group <*A.B.C.D* >**" command cancels the join.
ip igmp static-group <*A.B.C.D* > **no ip igmp static-group <*A.B.C.D*>**	Configure the interface to join in some IGMP static group; the "**no ip igmp static-group <*A.B.C.D* >**" command cancels the join.

（2）Configure IGMP Query parameters

1) Configure interval for IGMP to send query messages
2) Configure the maximum response time of IGMP query
3) Configure the time-out of IGMP query

Command	Explanation
Interface Configuration Mode	
ip igmp query-interval <*time_val*> **no ip igmp query-interval**	Configure the interval of IGMP query messages sent periodically; the "**no ip igmp query-interval**" command restores default value.
ip igmp query-max-response-time <*time_val*> **no ip igmp query-max-response-time**	Configure the maximum response time of the interface for IGMP query; the "**no ip igmp query-max-response-time**" command restores default value.
ip igmp query-timeout <*time_val*> **no ip igmp query-timeout**	Configure the time-out of the interface for IGMP query; the "**no ip igmp query-timeout**" command restores default value.

（3）Config IGMP version

Command	Explanation
Global Mode	
ip igmp version <*version*> **no ip igmp version**	Configure IGMP version on the interface; the "**no ip igmp version**" command restores the default value.

3. Disable IGMP Protocol

Command	Explanation
Interface Configuration Mode	
**no ip dvmrp	no ip pim dense-mode	no ip pim sparse-mode	no ip dvmrp multicast-routing	no ip pim multicast-routing**	Disable IGMP Protocol.

5.9.3 IGMP Configuration Examples

As shown in the following figure, add the Ethernet ports of Switch A and Switch B to

corresponding VLAN, and start PIM-DM on each VLAN interface.

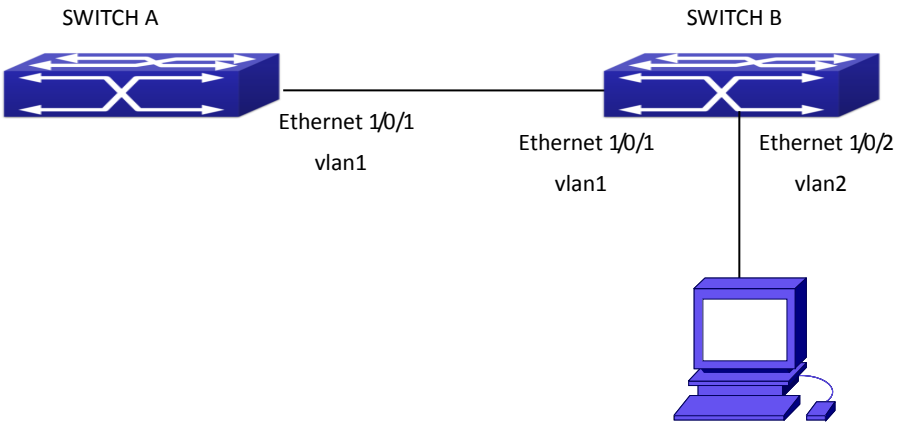


Figure 5-10 IGMP Network Topology Diagram

The configuration procedure for SwitchA and SwitchB is as follows:

(1) Configure SwitchA:

Switch(config)#ip pim multicast-routing

Switch (config)#interface vlan 1

Switch(Config-if-Vlan1)#ip address 12.1.1.1 255.255.255.0

Switch(Config-if-Vlan1)#ip pim dense-mode

(2) Configure SwitchB:

Switch(config)#ip pim multicast-routing

Switch(config)#interface vlan1

Switch(Config-if-Vlan1)#ip address 12.1.1.2 255.255.255.0

Switch(Config-if-Vlan1)#ip pim dense-mode

Switch(Config-if-Vlan1)#exit

Switch(config)#interface vlan2

Switch(Config-if-Vlan1)#ip address 20.1.1.1 255.255.255.0

Switch(Config-if-Vlan2)#ip pim dense-mode

Switch(Config-if-Vlan2)#ip igmp version 3

5.9.4 IGMP Troubleshooting

In configuring and using IGMP Protocol, IGMP Protocol might not operate normally caused by physical connection or incorrect configuration. Therefore, user should pay attention to the following issues:

- Firstly to assure that physical connection is correct;
- Next, to assure the Protocol of Interface and Link protocol is UP (use show interface command);
- Afterwards, to assure to start a kind of multicast protocol on the interface;
- Multicast Protocol requires RPF Check using unicast routing; therefore the correctness of unicast routing must be assured beforehand.

5.10 IGMP Snooping

5.10.1 Introduction to IGMP Snooping

IGMP (Internet Group Management Protocol) is a protocol used in IP multicast. IGMP is used by multicast enabled network device (such as a router) for host membership query, and by hosts that are joining a multicast group to inform the router to accept packets of a certain multicast address. All those operations are done through IGMP message exchange. The router will use a multicast address (224.0.0.1) that can address to all hosts to send an IGMP host membership query message. If a host wants to join a multicast group, it will reply to the multicast address of that a multicast group with an IGMP host membership reports a message.

IGMP Snooping is also referred to as IGMP listening. The switch prevents multicast traffic from flooding through IGMP Snooping, multicast traffic is forwarded to ports associated to multicast devices only. The switch listens to the IGMP messages between the multicast router and hosts, and maintains multicast group forwarding table based on the listening result, and can then decide to forward multicast packets according to the forwarding table.

Switch provides IGMP Snooping and is able to send a query from the switch so that the user can use switch in IP multicast.

5.10.2 IGMP Snooping Configuration Task List

1. Enable IGMP Snooping
2. Configure IGMP Snooping

1. Enable IGMP Snooping

Command	Explanation
Global Mode	
ip igmp snooping **no ip igmp snooping**	Enables IGMP Snooping. The no operation disables IGMP Snooping function.

2. Configure IGMP Snooping

Command	Explanation
Global Mode	
ip igmp snooping vlan *<vlan-id>* **no ip igmp snooping vlan *<vlan-id>***	Enables IGMP Snooping for specified VLAN. The no operation disables IGMP Snooping for specified VLAN.
ip igmp snooping proxy **no ip igmp snooping proxy**	Enable IGMP Snooping proxy function, the no command disables the function.

ip igmp snooping vlan < *vlan-id* > limit {group <*g_limit*> | source <*s_limit*>} **no ip igmp snooping vlan < *vlan-id* > limit**	Configure the max group count of vlan and the max source count of every group. The "no **ip igmp snooping vlan <*vlan-id*> limit"** command cancels this configuration.
ip igmp snooping vlan <1-4094> interface (ethernet | port-channel|) IFNAME limit {group <1-65535>| source <1-65535>} strategy (replace | drop) **no ip igmp snooping vlan <1-4094> interface (ethernet | port-channel|) IFNAME limit group source strategy**	Configure the number of groups which are allowed joining and the maximum of the source in each group under the IGMP Snooping port. Configure the strategy when it is up to the upper limit, including "replace" and "drop". No command configures as "no limitation".
ip igmp snooping vlan <*vlan-id*> l2-general-querier **no ip igmp snooping vlan <*vlan-id*> l2-general-querier**	Set this vlan to layer 2 general querier. It is recommended to configure a layer 2 general querier on a segment. The "no **ip igmp snooping vlan <*vlan-id*> l2-general-querier"command** cancels this configuration.
ip igmp snooping vlan <*vlan-id*> l2-general-querier-version <version>	Configure the version number of a general query from a layer 2 general querier.
ip igmp snooping vlan <*vlan-id*> l2-general-querier-source <source>	Configure the source address of a general query from a layer 2 general querier.
ip igmp snooping vlan <*vlan-id*> mrouter-port interface <*interface –name*> **no ip igmp snooping vlan <*vlan-id*> mrouter-port interface <*interface –name*>**	Configure static mrouter port of vlan. The no form of the command cancels this configuration.
ip igmp snooping vlan <*vlan-id*> mrouter-port learnpim **no ip igmp snooping vlan <*vlan-id*> mrouter-port learnpim**	Enable the function that the specified VLAN learns mrouter-port (according to pim packets), the no command will disable the function.
ip igmp snooping vlan <*vlan-id*> mrpt <*value* > **no ip igmp snooping vlan <*vlan-id*> mrpt**	Configure this survive time of mrouter port. The "no **ip igmp snooping vlan <*vlan-id*> mrpt"** command restores the default value.
ip igmp snooping vlan <*vlan-id*> query-interval <*value*> **no ip igmp snooping vlan <*vlan-id*> query-interval**	Configure this query interval. The "no **ip igmp snooping vlan <*vlan-id*> query-interval"** command restores the default value.

ip igmp snooping vlan *<vlan-id>* immediately-leave **no ip igmp snooping vlan *<vlan-id>* immediately-leave**	Enable the IGMP fast leave function for the specified VLAN: the "no **ip igmp snooping vlan *<vlan-id>* immediate-leave"** command disables the IGMP fast leave function.
ip igmp snooping vlan <*vlan-id*> query-mrsp <*value*> **no ip igmp snooping vlan <*vlan-id*> query-mrsp**	Configure the maximum query response period. The "no **ip igmp snooping vlan <*vlan-id*> query-mrsp"** command restores to the default value.
ip igmp snooping vlan *<vlan-id>* query-robustness <*value*> **no ip igmp snooping vlan *<vlan-id>* query-robustness**	Configure the query robustness. The "no **ip igmp snooping vlan *<vlan-id>* query-robustness"** command restores to the default value.
ip igmp snooping vlan <*vlan-id*> suppression-query-time <*value*> **no ip igmp snooping vlan <*vlan-id*> suppression-query-time**	Configure the suppression query time. The "no **ip igmp snooping vlan <*vlan-id*> suppression-query-time"** command restores to the default value.
ip igmp snooping vlan <*vlan-id*> static-group *<A.B.C.D>* [source *<A.B.C.D>*] interface [ethernet | port-channel] *<IFNAME>* **no ip igmp snooping vlan <*vlan-id*> static-group *<A.B.C.D>* [source *<A.B.C.D>*] interface [ethernet | port-channel] *<IFNAME>***	Configure static-group on specified port of the VLAN. The no form of the command cancels this configuration.
ip igmp snooping vlan <*vlan-id*> report source-address <*A.B.C.D*> **no ip igmp snooping vlan <*vlan-id*> report source-address**	Configure forwarding IGMP packet source address, The no operation cancels the packet source address.
ip igmp snooping vlan *<vlan-id>* specific-query-mrsp <value> **no ip igmp snooping vlan *<vlan-id>* specific-query-mrspt**	Configure the maximum query response time of the specific group or source, the no command restores the default value.
ip igmp snooping vlan <vlan-id> not-forward **no ip igmp snooping vlan <vlan-id> not-forward**	Configure the vlan not to forward IGMP packets to the mrouter port. The no command restores the default value.

5.10.3 IGMP Snooping Examples

Scenario 1: IGMP Snooping function

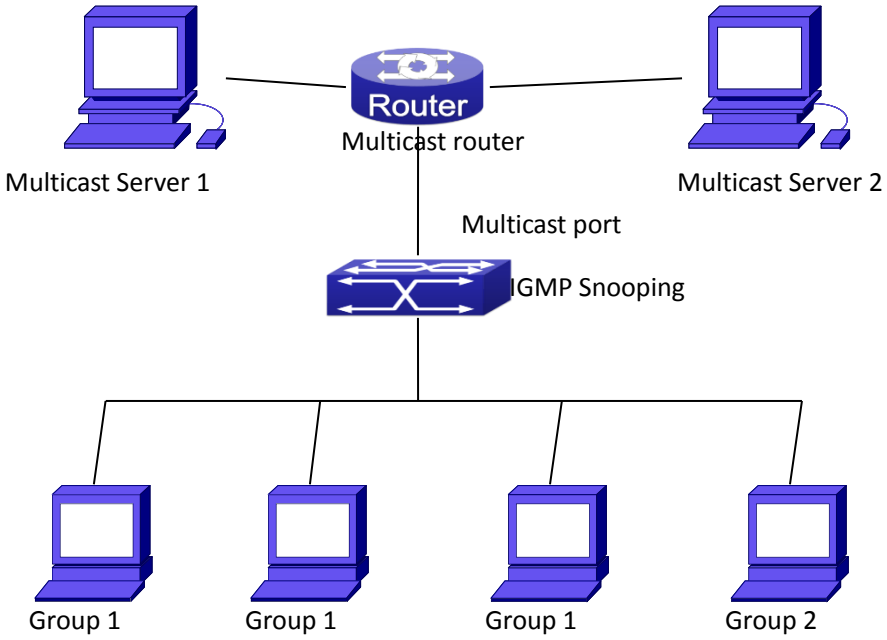


Figure 5-11 Enabling IGMP Snooping function

Example: As shown in the above figure, a VLAN 100 is configured in the switch and includes ports 1, 2, 6, 10 and 12. Four hosts are connected to port 2, 6, 10 and 12 respectively and the multicast router is connected to port 1. As IGMP Snooping is disabled by default either in the switch or in the VLANs, If IGMP Snooping should be enabled in VLAN 100, the IGMP Snooping should be first enabled for the switch in Global Mode and in VLAN 100 and set port 1 of VLAN 100 to be the mrouter port.

The configuration steps are listed below:
Switch(config)#ip igmp snooping
Switch(config)#ip igmp snooping vlan 100
Switch(config)#ip igmp snooping vlan 100 mrouter interface ethernet 1/0/1

Multicast Configuration

Suppose two programs are provided in the Multicast Server using multicast address Group1 and Group2, three of four hosts running multicast applications are connected to port 2, 6, 10 plays program1, while the host is connected to port 12 plays program 2.

IGMP Snooping listening result:

The multicast table built by IGMP Snooping in VLAN 100 indicates ports 1, 2, 6, 10 in Group1 and ports 1, 12 in Group2.

All the four hosts can receive the program of their choice: ports 2, 6, 10 will not receive the traffic of program 2 and port 12 will not receive the traffic of program 1.

Scenario 2: L2-general-querier

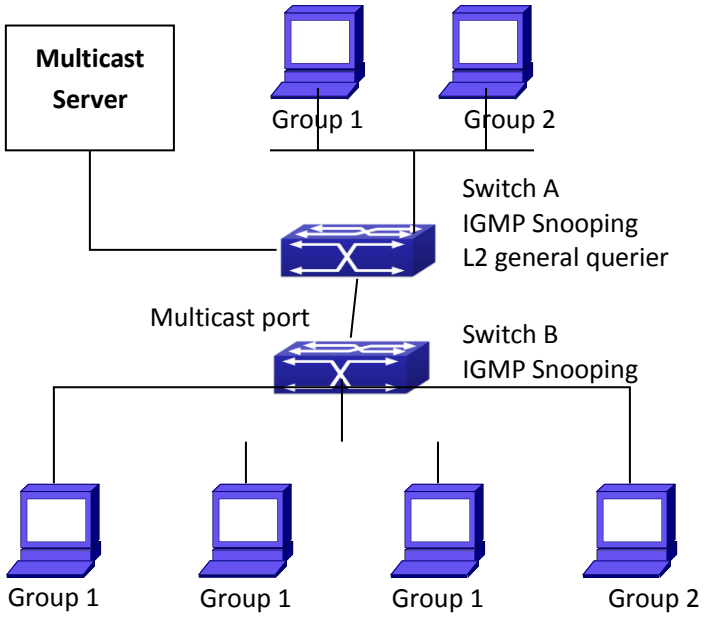


Figure 5-12 The switches as IGMP Queries

The configuration of Switch2 is the same as the switch in scenario 1, SwitchA takes the place of Multicast Router in scenario 1. Let's assume VLAN 60 is configured in SwitchA, including ports 1, 2, 10 and 12. Port 1 connects to the multicast server, and port 2 connects to Switch2. In order to send Query at regular interval, IGMP query must enabled in Global mode and in VLAN60.

The configuration steps are listed below:

SwitchA#config
SwitchA(config)#ip igmp snooping
SwitchA(config)#ip igmp snooping vlan 60
SwitchA(config)#ip igmp snooping vlan 60 L2-general-querier

SwitchB#config
SwitchB(config)#ip igmp snooping
SwitchB(config)#ip igmp snooping vlan 100
SwitchB(config)#ip igmp snooping vlan 100 mrouter interface ethernet 1/0/1

Multicast Configuration

The same as scenario 1

IGMP Snooping listening result:

Similar to scenario 1

Scenario 3: To run in cooperation with layer 3 multicast protocols.

SWITCH which is used in Scenario 1 is replaced with ROUTER with specific configurations remains the same. And multicast and IGMP snooping configurations are the same with what it is in Scenario 1. To configure PIM-SM on ROUTER, and enable PIM-SM on vlan 100 (use the same PIM mode with the connected multicast router)

Configurations are listed as below:
switch#config
switch(config)#ip pim multicast-routing
switch(config)#interface vlan 100
switch(config-if-vlan100)#ip pim sparse-mode

IGMP snooping does not distribute entries when layer 3 multicast protocol is enabled. It only does the following tasks.

- ☞ Remove the layer 2 multicast entries.
- ☞ Provide query functions to the layer 3 with vlan, S, and G as the parameters.
- ☞ When layer 3 IGMP is disabled, re-enable distributing layer 2 multicast entries.

By looking up the layer 3 IPMC entries, it can be found that ports can be indicated by the layer 3 multicast entries. This ensures the IGMP snooping can work in cooperation with the layer 3 multicast protocols.

5.10.4 IGMP Snooping Troubleshooting

On IGMP Snooping function configuration and usage, IGMP Snooping might not run properly because of physical connection or configuration mistakes. So the users should note that:

- ☞ Make sure correct physical connection
- ☞ Activate IGMP Snooping on whole configuration mode (use **ip igmp snooping**)
- ☞ Configure IGMP Snooping at VLAN on whole configuration mode (use **ip igmp snooping vlan <vlan-id>**)
- ☞ Make sure one VLAN is configured as L2 common checker in same mask, or make sure configured static mrouter
- ☞ Use **show ip igmp snooping vlan <vid>** command check IGMP Snooping information

5.11 IGMP Proxy

5.11.1 Introduction to IGMP Proxy

IGMP/MLD proxy which is introduced in rfc4605, is a simplified multicast protocol running at edge boxes. The edge boxes which runs the IGMP/MLD proxy protocol, does not need to run complicated multicast routing protocols such as PIM/DVMRP. However they work with multicast protocol enabled network through IGMP/MLD proxy. They can simplify the implementation of multicasting on edge devices.

The IGMP/MLD proxy works between the multicast router and the client, it works as both the multicast host and router. Upstream and downstream ports should be specified in the IGMP/MLD proxy configuration. The host protocol runs at upstream ports, while the router protocol runs at downstream ports. The switch collects the join and leave messages received from downstream ports and forward them to the multicast router through upstream ports.

The IGMP proxy configuration is exclusive with PIM and DVMRP configuration.

5.11.2 IGMP Proxy Configuration Task List

1. Enable IGMP Proxy function
2. Enable configurations for both downstream and upstream ports for the IGMP Proxy in different interfaces
3. Configure IGMP Proxy

1. Enable IGMP Proxy function

Command	Explanation
Global Mode	
ip igmp proxy **no ip igmp proxy**	Enable IGMP Proxy function. The "**no ip igmp proxy**" disables this function.

2. Enable configurations for both downstream and upstream ports for the IGMP Proxy in different interfaces

Command	Explanation
Interface Configuration Mode	
ip igmp proxy upstream **no ip igmp proxy upstream**	Enable IGMP Proxy upstream function. The "**no ip igmp proxy upstream**" disables this function.
ip igmp proxy downstream **no ip igmp proxy downstream**	Enable IGMP Proxy downstream function. The "**no ip igmp proxy downstream**" disables this function.

3. Configure IGMP Proxy assistant parameter

Command	Explanation
Global Mode	
ip igmp proxy limit {group *<1-500>*| source *<1-500>*} **no ip igmp proxy limit**	To configure the maximum number of groups that upstream ports can join, and the maximum number of sources in a single group. The no form of this command will restore the default value.
ip igmp proxy unsolicited-report interval *<1-5>* **no ip igmp proxy unsolicited-report interval**	To configure how often the upstream ports send out unsolicited report. The no form of this command will restore the default configuration.
ip igmp proxy unsolicited-report robustness	To configure the retry times of upstream ports'

<2-10> **no ip igmp proxy unsolicited-report robustness**	sending unsolicited reports. The no form of this command will restore the default value.
ip igmp proxy aggregate **no ip igmp proxy aggregate**	To configure non-query downstream ports to be able to aggregate the IGMP operations. The no form of this command will restore the default configuration.
ip multicast ssm range *<1-99>* **ip multicast ssm default** **no ip mulitcast ssm**	To configure the address range for IGMP proxy ssm multicast groups; The no form of this command will remove the configuration.
ip igmp proxy multicast-source **no ip igmp proxy multicast-source**	To configure the port as downstream ports for the source of multicast datagram; The no from of this command will disable the configuration.

5.11.3 IGMP Proxy Examples

Example 1: IGMP Proxy function.

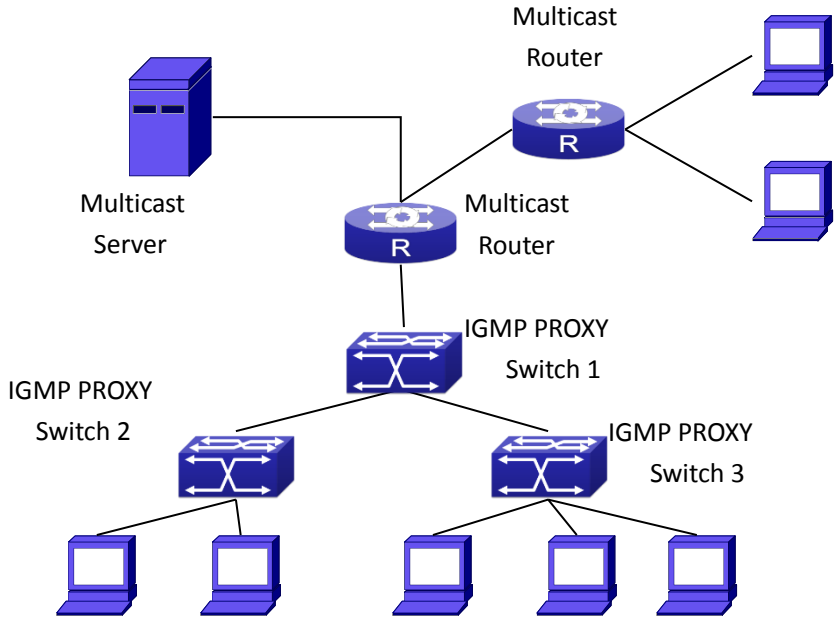


Figure 5-13 IGMP Proxy Topology Diagram

As it is show in the figure above, the switch functions as IGMP Proxy in a network of topology of tree, the switch aggregates the multicast dataflow from upstream port and redistributes them to the downstream ports, while the IGMP membership reports flow from downstream ports to upstream ports. Three IGMP Proxy enabled switches which are connected in tree topology, respectively have one port connected to multicast routers, and no less than one

ports connected to hosts or upstream ports from other IGMP Proxy enabled switches.

The configuration steps are listed below:

Switch#config
Switch(config)#ip igmp proxy
Switch(Config)#interface vlan 1
Switch(Config-if-Vlan1)#ip igmp proxy upstream
Switch(Config)#interface vlan 2
Switch(Config-if-Vlan2)#ip igmp proxy downstream

Multicast Configuration:

Suppose the multicast server offers some programs through 224.1.1.1. Some hosts subscribe that program at the edge of the network. The IGMP multicast members report themselves to the downstream ports of IGMP Proxy enabled Switch 2 and Switch 3. Switch 2 and Switch 3 then aggregate the group membership information and send them through the upstream ports. Switch 1 finally forward these membership information to the multicast router when receiving the group membership information through upstream ports, and deliver the multicast dataflow through downstream ports.

Example2: IGMP Proxy for multicast sources from downstream ports.

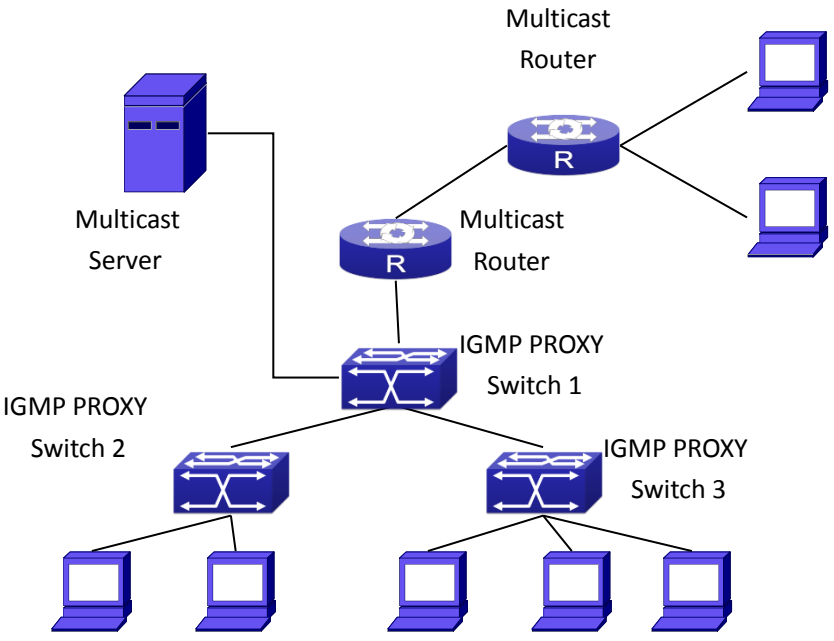


Figure 5-14 IGMP Proxy for multicast sources from downstream ports

As it is show in the figure above, IGMP Proxy enabled switches connected to the network in tree topology. The multicast source server connects to the downstream port of Switch1, the multicast dataflow is distributed through the upstream port and other downstream ports. Three IGMP Proxy enabled switches which are connected in tree topology, respectively have one port connected to multicast routers, and no less than one ports connected to hosts or upstream ports

from other IGMP proxy enabled switches.

The configuration steps are listed below:

IGMP PROXY Switch1 configuration：

Switch#config

Switch(config)#ip igmp proxy

Switch(Config)#interface vlan 1

Switch(Config-if-Vlan1)#ip igmp proxy upstream

Switch(Config)#interface vlan 2

Switch(Config-if-Vlan2)#ip igmp proxy downstream

Switch(Config-if-Vlan2)#ip igmp proxy multicast-source

Route1 configuration:

Switch#config

Switch(config)#ip pim multicast

Switch(Config)#interface vlan 1

Switch(Config-if-Vlan1)#ip pim sparse-mode

Switch(Config-if-Vlan1)#ip pim bsr-border

Multicast Configuration:

Suppose the server provides programs through the multicast address 224.1.1.1, and some hosts subscribe that program on the edge of the network. The host reports their IGMP multicast group membership to Switch 2 and Switch 3 through downstream ports. Switch 2 and Switch 3 then aggregate and forward them to Switch 1 which then forwards the information to multicast router. When multicast dataflow arrives, the IGMP Proxy enabled switches re-distribute the group membership through upstream ports and downstream ports. When the multicast router receives the multicast dataflow from IGMP proxy, it will consider the multicast data source is directly connected to the router, and determine the identity of DR and ORIGINATOR. The multicast dataflow will be redistributed according to the PIM protocol.

5.11.4 IGMP Proxy Troubleshooting

When IGMP Proxy function configuration and usage, IGMP Proxy might not run properly because of physical connection or configuration mistakes. So the users should note that:

- ☞ Make sure physical connection correctly;
- ☞ Activate IGMP Proxy on whole Global mode (use **ip igmp proxy**);
- ☞ Make sure configure one upstream port and at least one downstream port under interface configuration mode (Use **ip igmp proxy upstream**, **ip igmp proxy downstream**);
- ☞ Use **show ip igmp proxy** command to check if the IGMP Proxy information is correct.

If the IGMP Proxy problem remains unsolved, please use debug IGMP Proxy and other debugging command and copy the DEBUG message within three minutes, send the recorded message to the technical service center of our company.

5.12 Multicast VLAN

5.12.1 Introductions to Multicast VLAN

Based on current multicast order method, when orders from users in different VLAN, each VLAN will copy a multicast traffic in this VLAN, which is a great waste of the bandwidth. By configuration of the multicast VLAN, we add the switch port to the multicast VLAN, with the IGMP Snooping/MLD Snooping functions enabled, users from different VLAN will share the same multicast VLAN. The multicast traffic only exists within a multicast VLAN, so the bandwidth is saved. As the multicast VLAN is absolutely separated from the user VLAN, security and bandwidth concerns can be met at the same time, after the multicast VLAN is configured, the multicast traffic will be continuously sent to the users.

5.12.2 Multicast VLAN Configuration Task List

1. Enable the multicast VLAN function
2. Configure the IGMP Snooping
3. Configure the MLD Snooping

1. Enable the multicast VLAN function

Command	Explanation
VLAN configuration mode	
multicast-vlan **no multicast-vlan**	Configure a VLAN and enable the multicast VLAN on it. The "**no multicast-vlan**" command disables the multicast function on the VLAN.
multicast-vlan association *<vlan-list>* **no multicast-vlan association *<vlan-list>***	Associate a multicast VLAN with several VLANs. The no form of this command deletes the related VLANs associated with the multicast VLAN.
multicast-vlan association interface (ethernet | port-channel|) IFNAME **no multicast-vlan association interface (ethernet | port-channel|) IFNAME**	Associate the specified port with the multicast VLAN, so the associated ports are able to receive the multicast flow. The no command cancels the association between the ports and the multicast VLAN.
multicast-vlan mode {dynamic| compatible} **no multicast-vlan mode {dynamic| compatible}**	Configure the two modes of multicast vlan. The no command cancels the mode configuration.

2. Configure the IGMP Snooping

Command	Explanation
Global Mode	
ip igmp snooping vlan *<vlan-id>* **no ip igmp snooping vlan *<vlan-id>***	Enable the IGMP Snooping function on the multicast VLAN. The no form of this command disables the IGMP Snooping on the multicast VLAN.
ip igmp snooping **no ip igmp snooping**	Enable the IGMP Snooping function. The no form of this command disables the IGMP snooping function.

3. Configure the MLD Snooping

ipv6 mld snooping vlan *<vlan-id>* **no ipv6 mld snooping vlan *<vlan-id>***	Enable MLD Snooping on multicast VLAN; the no form of this command disables MLD Snooping on multicast VLAN.
ipv6 mld snooping **no ipv6 mld snooping**	Enable the MLD Snooping function. The no form of this command disables the MLD snooping function.

5.12.3 Multicast VLAN Examples

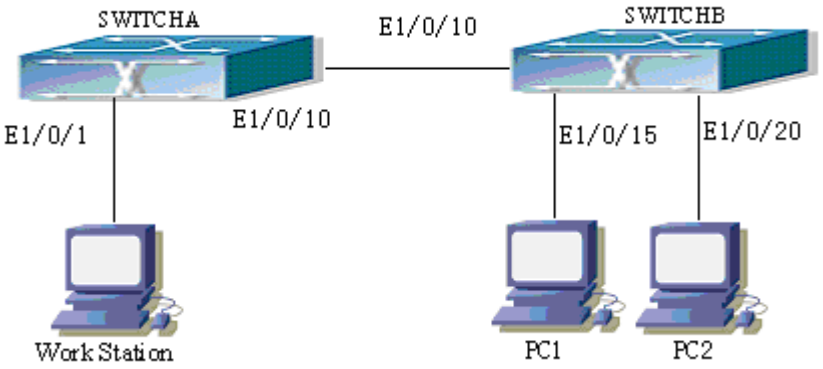


Figure 5-15 Function configuration of the Multicast VLAN

As shown in the figure, the multicast server is connected to the layer 3 switch switchA through port 1/0/1 which belongs to the VLAN10 of the switch. The layer 3 switch switchA is connected with layer 2 switches through the port1/0/10, which configured as trunk port. On the switchB the VLAN100 is configured set to contain port1/0/15, and VLAN101 to contain port1/0/20. PC1 and PC2 are respectively connected to port 1/0/15 and1/0/20. The switchB is connected with the switchA through port1/0/10, which configured as trunk port. VLAN 20 is a multicast VLAN. By configuring multicast vlan, the PC1 and PC2 will receives the multicast data from the multicast VLAN.

Following configuration is based on the IP address of the switch has been configured and all the equipment are connected correctly.

Configuration procedure

SwitchA#config

```
SwitchA(config)#vlan 10
SwitchA(config-vlan10)#switchport access ethernet 1/0/1
SwitchA(config-vlan10)exit
SwitchA(config)#interface vlan 10
Switch(Config-if-Vlan10)#ip pim dense-mode
Switch(Config-if-Vlan10)#exit
SwitchA(config)#vlan 20
SwitchA(config-vlan20)#exit
SwitchA(config)#interface vlan 20
SwitchA(Config-if-Vlan20)#ip pim dense-mode
SwitchA(Config-if-Vlan20)#exit
SwitchA(config)#ip pim multicast
SwitchA(config)# interface ethernet1/0/10
SwitchA(Config-If-Ethernet1/0/10)switchport mode trunk

SwitchB#config
SwitchB(config)#vlan 100
SwitchB(config-vlan100)#Switchport access ethernet 1/0/15
SwitchB(config-vlan100)exit
SwitchB(config)#vlan 101
SwitchB(config-vlan101)#Switchport access ethernet 1/0/20
SwitchB(config-vlan101)exit
SwitchB(config)# interface ethernet 1/0/10
SwitchB(Config-If-Ethernet1/0/10)#switchport mode trunk
SwitchB(Config-If-Ethernet1/0/10)#exit
SwitchB(config)#vlan 20
SwitchB(config-vlan20)#multicast-vlan
SwitchB(config-vlan20)#multicast-vlan association 100,101
SwitchB(config-vlan20)#exit
SwitchB(config)#ip igmp snooping
SwitchB(config)#ip igmp snooping vlan 20
```

When multicast VLAN supports IPv6 multicast, usage is the same with IPv4, but the difference is using with MLD Snooping, so does not give an example.

Chapter 6 Security Function Configuration

6.1 ACL

6.1.1 Introduction to ACL

ACL (Access Control List) is an IP packet filtering mechanism employed in switches, providing network traffic control by granting or denying access the switches, effectively safeguarding the security of networks. The user can lay down a set of rules according to some information specific to packets, each rule describes the action for a packet with certain information matched: 'permit' or 'deny'. The user can apply such rules to the incoming direction of switch ports, so that data streams in the incoming direction of specified ports must comply with the ACL rules assigned.

Egress ACL can implement the filtering of the packets on egress and ingress direction, the packets match the specific rules can be allowed or denied. ACL can support IP ACL, MAC ACL, MAC-IP ACL, IPv6 ACL. Ingress direction of the port can bind four kinds of ACL at the same time, there are four resources on egress direction of the port, IP ACL and MAC ACL engage one resource severally, MAC-IP ACL and IPv6 ACL engage two resources severally, so egress direction of the port can not bind four kinds of ACL at the same time. When binding three kinds of ACL at the same time, it should be the types of IP, MAC, MAC-IP or IP, MAC, IPv6. When binding two kinds of ACL at the same time, any combination of ACL type is valid. Each type can only apply one on the port.

6.1.1.1 Access-list

Access-list is a sequential collection of conditions that corresponds to a specific rule. Each rule consist of filter information and the action when the rule is matched. Information included in a rule is the effective combination of conditions such as source IP, destination IP, IP protocol number and TCP port, UDP port. Access-lists can be categorized by the following criteria:

- ☞ Filter information based criterion: IP access-list (layer 3 or higher information), MAC access-list (layer 2 information), and MAC-IP access-list (layer 2 or layer 3 or higher).
- ☞ Configuration complexity based criterion: standard and extended, the extended mode allows more specific filtering of information.
- ☞ Nomenclature based criterion: numbered and named.

Description of an ACL should cover the above three aspects.

6.1.1.2 Access-group

When a set of access-lists are created, they can be applied to traffic of incoming direction on all ports. Access-group is the description to the binding of an access-list to the incoming direction on a specific port. When an access-group is created, all packets from in the incoming direction through the port will be compared to the access-list rule to decide whether to permit or deny access.

The current firmware only supports ingress ACL configuration.

6.1.1.3 Access-list Action and Global Default Action

There are two access-list actions and default actions: “permit” or “deny”. The following rules apply:

- ☞ An access-list can consist of several rules. Filtering of packets compares packet conditions to the rules, from the first rule to the first matched rule; the rest of the rules will not be processed.
- ☞ Global default action applies only when packet flirter is enabled on a port and no ACL is bound to that port, or no binding ACL matches.

6.1.2 ACL Configuration Task List

ACL Configuration Task Sequence:

1. Configuring access-list
   (1) Configuring a numbered standard IP access-list
   (2) Configuring a numbered extended IP access-list
   (3) Configuring a standard IP access-list based on nomenclature
      a) Create a standard IP access-list based on nomenclature
      b) Specify multiple “permit” or “deny” rule entries
      c) Exit ACL Configuration Mode
   (4) Configuring an extended IP access-list based on nomenclature
      a) Create an extensive IP access-list based on nomenclature
      b) Specify multiple “permit” or “deny” rule entries
      c) Exit ACL Configuration Mode
   (5) Configuring a numbered standard MAC access-list
   (6) Configuring a numbered extended MAC access-list
   (7) Configuring a extended MAC access-list based on nomenclature
      a) Create a extensive MAC access-list based on nomenclature
      b) Specify multiple “permit” or “deny” rule entries
      c) Exit ACL Configuration Mode
   (8) Configuring a numbered extended MAC-IP access-list
   (9) Configuring a extended MAC-IP access-list based on nomenclature

a) Create a extensive MAC-IP access-list based on nomenclature

b) Specify multiple “permit” or “deny” rule entries

c) Exit MAC-IP Configuration Mode

(10) Configuring a numbered standard IPv6 access-list

(11) Configuring a numbered extended IPv6 access-list

(12) Configuring a standard IPv6 access-list based on nomenclature

a) Create a standard IPv6 access-list based on nomenclature

b) Specify multiple permit or deny rule entries

c) Exit ACL Configuration Mode

(13) Configuring an extended IPv6 access-list based on nomenclature.

a) Create an extensive IPv6 access-list based on nomenclature

b) Specify multiple permit or deny rule entries

c) Exit ACL Configuration Mode

2. Configuring the packet filtering function

(1) Enable global packet filtering function

(2) Configure ACL deny preemption function globally (optional)

3. Configuring time range function

(1) Create the name of the time range

(2) Configure periodic time range

(3) Configure absolute time range

4. Bind access-list to an incoming direction of the specified port

5. Show configuration and statistic information of ACL

6. Clear the filtering information of the specified port

1. Configuring access-list

(1) Configuring a numbered standard IP access-list

Command	Explanation
Global Mode	
access-list *<num>* {deny | permit} {{*<sIpAddr>* *<sMask>*} | any-source | {host-source *<sIpAddr>*}} **no access-list *<num>***	Creates a numbered standard IP access-list, if the access-list already exists, then a rule will add to the current access-list; the “**no access-list *<num>***“ command deletes a numbered standard IP access-list.

(2) Configuring a numbered extensive IP access-list

Command	Explanation
Global Mode	

access-list *<num>* {deny | permit} icmp {{*<sIpAddr>* *<sMask>*} | any-source | {host-source *<sIpAddr>*}} {{*<dIpAddr>* *<dMask>*} | any-destination | {host-destination *<dIpAddr>*}} [*<icmp-type>* [*<icmp-code>*]] [precedence *<prec>*] [tos *<tos>*][time-range*<time-range-name>*]	Creates a numbered ICMP extended IP access rule; if the numbered extended access-list of specified number does not exist, then an access-list will be created using this number.
access-list *<num>* {deny | permit} igmp {{*<sIpAddr>* *<sMask>*} | any-source | {host-source *<sIpAddr>*}} {{*<dIpAddr>* *<dMask>*} | any-destination | {host-destination *<dIpAddr>*}} [*<igmp-type>*] [precedence *<prec>*] [tos *<tos>*][time-range*<time-range-name>*]	Creates a numbered IGMP extended IP access rule; if the numbered extended access-list of specified number does not exist, then an access-list will be created using this number.
access-list *<num>* {deny | permit} tcp {{*<sIpAddr>* *<sMask>*} | any-source | {host-source *<sIpAddr>*}} [s-port {*<sPort>* | range *<sPortMin>* *<sPortMax>*}] {{*<dIpAddr>* *<dMask>*} | any-destination | {host-destination *<dIpAddr>*}} [d-port {*<dPort>* | range *<dPortMin>* *<dPortMax>*}] [ack+fin+psh+rst+urg+syn] [precedence *<prec>*] [tos *<tos>*][time-range*<time-range-name>*]	Creates a numbered TCP extended IP access rule; if the numbered extended access-list of specified number does not exist, then an access-list will be created using this number.
access-list *<num>* {deny | permit} udp {{*<sIpAddr>* *<sMask>*} | any-source | {host-source *<sIpAddr>*}} [s-port {*<sPort>* | range *<sPortMin>* *<sPortMax>*}] {{*<dIpAddr>* *<dMask>*} | any-destination | {host-destination *<dIpAddr>*}} [d-port {*<dPort>* | range *<dPortMin>* *<dPortMax>*}] [precedence *<prec>*] [tos *<tos>*][time-range*<time-range-name>*]	Creates a numbered UDP extended IP access rule; if the numbered extended access-list of specified number does not exist, then an access-list will be created using this number.
access-list *<num>* {deny | permit} {eigrp | gre | igrp | ipinip | ip | ospf | *<protocol-num>*} {{*<sIpAddr>* *<sMask>*} | any-source | {host-source *<sIpAddr>*}} {{*<dIpAddr>* *<dMask>*} | any-destination | {host-destination *<dIpAddr>*}} [precedence *<prec>*] [tos *<tos>*][time-range*<time-range-name>*]	Creates a numbered IP extended IP access rule for other specific IP protocol or all IP protocols; if the numbered extended access-list of specified number does not exist, then an access-list will be created using this number.
no access-list *<num>*	Deletes a numbered extensive IP access-list.

(3) Configuring a standard IP access-list basing on nomenclature

a. Create a name-based standard IP access-list

Command	Explanation
Global Mode	

ip access-list standard <*name*> **no ip access-list standard <*name*>**	Creates a standard IP access-list based on nomenclature; the "**no ip access-list standard <*name*>**" command deletes the name-based standard IP access-list.

b. Specify multiple "permit" or "deny" rules

Command	Explanation
Standard IP ACL Mode	
[no] {deny | permit} {{<*sIpAddr*> <*sMask*>} | any-source | {host-source <*sIpAddr*>}}	Creates a standard name-based IP access rule; the "**no**" form command deletes the name-based standard IP access rule.

c. Exit name-based standard IP ACL configuration mode

Command	Explanation
Standard IP ACL Mode	
exit	Exits name-based standard IP ACL configuration mode.

(4) Configuring an name-based extended IP access-list

a. Create an extended IP access-list basing on nomenclature

Command	Explanation
Global Mode	
ip access-list extended <*name*> **no ip access-list extended <*name*>**	Creates an extended IP access-list basing on nomenclature; the "**no ip access-list extended <name>** " command deletes the name-based extended IP access-list.

b. Specify multiple "permit" or "deny" rules

Command	Explanation
Extended IP ACL Mode	
[no] {deny | permit} icmp {{<*sIpAddr*> <*sMask*>} | any-source | {host-source <*sIpAddr*>}} {{<*dIpAddr*> <*dMask*>} | any-destination | {host-destination <*dIpAddr*>}} [<*icmp-type*> [<*icmp-code*>]] [precedence <*prec*>] [tos <*tos*>][time-range<*time-range-name*>]	Creates an extended name-based ICMP IP access rule; the no form command deletes this name-based extended IP access rule.
[no] {deny | permit} igmp {{<*sIpAddr*> <*sMask*>} | any-source | {host-source <*sIpAddr*>}} {{<*dIpAddr*> <*dMask*>} | any-destination | {host-destination <*dIpAddr*>}} [<*igmp-type*>] [precedence <*prec*>] [tos <*tos*>][time-range<*time-range-name*>]	Creates an extended name-based IGMP IP access rule; the no form command deletes this name-based extended IP access rule.

[no] {deny | permit} tcp {{*<sIpAddr> <sMask>*} | any-source | {host-source *<sIpAddr>*}} [s-port {*<sPort>* | range *<sPortMin> <sPortMax>*}] {{*<dIpAddr> <dMask>*} | any-destination | {host-destination *<dIpAddr>*}} [d-port {*<dPort>* | range *<dPortMin> <dPortMax>*}] [ack+fin+psh+rst+urg+syn] [precedence *<prec>*] [tos *<tos>*][time-range*<time-range-name>*]	Creates an extended name-based TCP IP access rule; the no form command deletes this name-based extended IP access rule.
[no] {deny | permit} udp {{*<sIpAddr> <sMask>*} | any-source | {host-source *<sIpAddr>*}} [s-port {*<sPort>* | range *<sPortMin> <sPortMax>*}] {{*<dIpAddr> <dMask>*} | any-destination | {host-destination *<dIpAddr>*}} [d-port {*<dPort>* | range *<dPortMin> <dPortMax>*}] [precedence *<prec>*] [tos *<tos>*][time-range*<time-range-name>*]	Creates an extended name-based UDP IP access rule; the no form command deletes this name-based extended IP access rule.
[no] {deny | permit} {eigrp | gre | igrp | ipinip | ip | ospf | <protocol-num>} {{*<sIpAddr> <sMask>*} | any-source | {host-source *<sIpAddr>*}} {{*<dIpAddr> <dMask>*} | any-destination | {host-destination *<dIpAddr>*}} [precedence *<prec>*] [tos *<tos>*][time-range*<time-range-name>*]	Creates an extended name-based IP access rule for other IP protocols; the no form command deletes this name-based extended IP access rule.

c. Exit extended IP ACL configuration mode

Command	Explanation
Extended IP ACL Mode	
exit	Exits extended name-based IP ACL configuration mode.

(5) Configuring a numbered standard MAC access-list

Command	Explanation
Global Mode	
access-list*<num>*{deny|permit}{any-source-mac|{host-source-mac*<host_smac>*}|{*<smac><smac-mask>*}} **no access-list *<num>***	Creates a numbered standard MAC access-list, if the access-list already exists, then a rule will add to the current access-list; the "**no access-list *<num>***" command deletes a numbered standard MAC access-list.

(6) Creates a numbered MAC extended access-list

Command	Explanation
Global Mode	
access-list*<num>* {deny|permit} {any-source-mac| {host-source-mac*<host_smac>*}|{*<smac><smac-mask>*}}{ any-destination-mac|{host-destination-mac*<host_dmac>* }|{*<dmac><dmac-mask>*}}[{untagged-eth2 | tagged-eth2 | untagged-802-3 | tagged-802-3} [*<offset1> <length1>*	Creates a numbered MAC extended access-list, if the access-list already exists, then a rule will add to the current access-list; the "**no access-list**

<value1> [*<offset2>* *<length2>* <value2> [<offset3> <length3> <value3> [*<offset4>* *<length4>* *<value4>*]]]] **no access-list *<num>***	***<num>*“ command deletes a numbered MAC extended access-list.

(7) Configuring a extended MAC access-list based on nomenclature

a. Create an extensive MAC access-list based on nomenclature

Command	Explanation
Global Mode	
mac-access-list extended *<name>* **no mac-access-list extended *<name>***	Creates an extended name-based MAC access rule for other IP protocols; the no form command deletes this name-based extended MAC access rule.

b. Specify multiple “permit” or “deny” rule entries

Command	Explanation
Extended name-based MAC access rule Mode	
[no]{deny|permit}{any-source-mac|{host-source-mac*<host_smac>*}|{*<smac>*<*smac-mask*>}} **{any-destination-mac|{host-destination-mac *<host_dmac>*} |{*<dmac>* *<dmac-mask>*}} [cos *<cos-val>* [*<cos-bitmask>*] [vlanId *<vid-value>* [*<vid-mask>*][ethertype*<protocol>*[*<protocol-mask>*]]]]** **[no]{deny|permit} {any-source-mac|{host-source-mac*<host_smac>*}|{*<smac>*<*smac-mask*>}}{any-destination-mac|{host-destination-mac*<host_dmac>*}|{*<dmac>*<*dmac-mask*>}} [ethertype *<protocol>* [*<protocol-mask>*]]** **[no]{deny|permit}{any-source-mac|{host-source-mac*<host_smac>*}|{*<smac>*<*smac-mask*>}} {any-destination-mac|{host-destination-mac*<host_dmac>*}|{*<dmac>*<*dmac-mask*>}} [vlanid *<vid-value>* [*<vid-mask>*][ethertype *<protocol>* [*<protocol-mask>*]]]**	Creates an extended name-based MAC access rule matching MAC frame; the no form command deletes this name-based extended MAC access rule.
[no]{deny|permit}{any-source-mac|{host-source-mac*<host_smac>*}|{*<smac>*<*smac-mask*>}}{any-destination-mac|{host-destination-mac*<host_dmac>*}|{*<dmac>*<*dmac-mask*>}}[untagged-eth2 [ethertype *<protocol>* [protocol-mask]]]	Creates an extended name-based MAC access rule matching untagged ethernet 2 frame; the no form command deletes this name-based extended MAC access rule.

[no]{deny|permit}{any-source-mac|{host-source-mac<*host_smac*>}|{<*smac*><*smac-mask*>}} {any-destination-mac|{host-destination-mac <*host_dmac*>}|{<*dmac*><*dmac-mask*>}} [untagged-802-3]	Creates an name-based extended MAC access rule matching 802.3 frame; the no form command deletes this name-based extended MAC access rule.
[no]{deny|permit}{any-source-mac|{host-source-mac<*host_smac*>}|{<*smac*><*smac-mask*>}}{any-destination-mac|{host-destination-mac<*host_dmac*>}|{<*dmac*><*dmac-mask*>}}[tagged-eth2 [cos <*cos-val*> [<*cos-bitmask*>]] [vlanId <*vid-value*> [<*vid-mask*>]] [ethertype<*protocol*> [<*protocol-mask*>]]]	Creates an name-based extended MAC access rule matching tagged ethernet 2 frame; the no form command deletes this name-based extended MAC access rule.
[no]{deny|permit}{any-source-mac|{host-source-mac <*host_smac*>}|{<*smac*><*smac-mask*>}} {any-destination-mac|{host-destination-mac<*host_dmac*>}|{<*dmac*><*dmac-mask*>}} [tagged-802-3 [cos <*cos-val*> [<*cos-bitmask*>]] [vlanId <*vid-value*> [<*vid-mask*>]]]	Creates an name-based extended MAC access rule matching tagged 802.3 frame; the no form command deletes this name-based extended MAC access rule.

c. Exit ACL Configuration Mode

Command	Explanation
Extended name-based MAC access configure Mode	
exit	Quit the extended name-based MAC access configure mode.

(8) Configuring a numbered extended MAC-IP access-list

Command	Explanation
Global mode	
access-list<*num*>{deny|permit} {any-source-mac| {host-source-mac <*host_smac*>} | {<*smac*> <*smac-mask*>}} {any-destination-mac | {host-destination-mac <*host_dmac*>} | {<*dmac*><*dmac-mask*>}} icmp {{<*source*> <*source-wildcard*>} |any-source| {host-source <*source-host-ip*>}} {{<*destination*> <*destination-wildcard*>} | any-destination | {host-destination <*destination-host-ip*>}} [<*icmp-type*> [<*icmp-code*>]] [precedence <*precedence*>] [tos <*tos*>] [time-range <*time-range-name*>]	Creates a numbered mac-icmp extended mac-ip access rule; if the numbered extended access-list of specified number does not exist, then an access-list will be created using this number.
access-list<*num*>{deny|permit}{any-source-mac| {host-source-mac<*host_smac*>}|{<*smac*><*smac-mask*>}} {any-destination-mac|{host-destination-mac <*host_dmac*>}|{<*dmac*><*dmac-mask*>}}igmp {{<*source*><*source-wildcard*>}|any-source|	Creates a numbered mac-igmp extended mac-ip access rule; if the numbered extended access-list of specified number does not exist, then an

{host-source<*source-host-ip*>} **{{<*destination*><*destination-wildcard*>}|any-destination|** **{host-destination<*destination-host-ip*>}}** **[<*igmp-type*>] [precedence <*precedence*>] [tos <*tos*>][time-range<*time-range-name*>]**	access-list will be created using this number.
access-list<num>{deny|permit}{any-source-mac| **{host-source-mac<host_smac>}|{<smac><smac-mask>}}{** **any-destination-mac|{host-destination-mac** **<host_dmac>}|{<dmac><dmac-mask>}}tcp** **{{<source><source-wildcard>}|any-source|** **{host-source<source-host-ip>}} [s-port {<*port1*> | range <*sPortMin*> <*sPortMax*>}]** **{{<destination><destination-wildcard>}|any-destination|** **{host-destination <destination-host-ip>}} [d-port {<*port3*> | range <*dPortMin*> <*dPortMax*>}]** **[ack+fin+psh+rst+urg+syn] [precedence <precedence>]** **[tos <tos>][time-range<*time-range-name*>]**	Creates a numbered mac-ip extended mac-tcp access rule; if the numbered extended access-list of specified number does not exist, then an access-list will be created using this number.
access-list<*num*>{deny|permit}{any-source-mac| **{host-source-mac<*host_smac*>}|{<*smac*><*smac-mask*>}}{** **any-destination-mac|{host-destination-mac** **<*host_dmac*>}|{<*dmac*><*dmac-mask*>}}udp** **{{<*source*><*source-wildcard*>}|any-source|** **{host-source<*source-host-ip*>}} [s-port {<*port1*> | range <*sPortMin*> <*sPortMax*>}]** **{{<*destination*><*destination-wildcard*>}|any-destination|** **{host-destination<*destination-host-ip*>}} [d-port {<*port3*> | range <*dPortMin*> <*dPortMax*>}] [precedence <*precedence*>] [tos <*tos*>][time-range<*time-range-name*>]**	Creates a numbered mac-udp extended mac-ip access rule; if the numbered extended access-list of specified number does not exist, then an access-list will be created using this number.
access-list<*num*>{deny|permit}{any-source-mac| **{host-source-mac<*host_smac*>}|{<*smac*><*smac-mask*>}}** **{any-destination-mac|{host-destination-mac** **<*host_dmac*>}|{<*dmac*><*dmac-mask*>}}** **{eigrp|gre|igrp|ip|ipinip|ospf|{<*protocol-num*>}}** **{{<*source*><*source-wildcard*>}|any-source|** **{host-source<*source-host-ip*>}}** **{{<*destination*><*destination-wildcard*>}|any-destination|** **{host-destination<*destination-host-ip*>}}** **[precedence <*precedence*>] [tos <*tos*>][time-range<*time-range-name*>]**	Creates a numbered extended mac-ip access rule for other specific mac-ip protocol or all mac-ip protocols; if the numbered extended access-list of specified number does not exist, then an access-list will be created using this number.
no access-list <*num*>	Deletes this numbered extended MAC-IP access rule.

(9) Configuring a extended MAC-IP access-list based on nomenclature

a. Create an extensive MAC-IP access-list based on nomenclature

Command	Explanation
Global Mode	
mac-ip-access-list extended *<name>* **no mac-ip-access-list extended *<name>***	Creates an extended name-based MAC-IP access rule; the no form command deletes this name-based extended MAC-IP access rule.

b. Specify multiple "permit" or "deny" rule entries

Command	Explanation
Extended name-based MAC-IP access Mode	
[no]{deny|permit} {any-source-mac|{host-source-mac *<host_smac>*}|{*<smac>*<*smac-mask>*}} {any-destination-mac|{host-destination-mac *<host_dmac>*}|{*<dmac>*<*dmac-mask>*}}icmp {{*<source>*<*source-wildcard>*}|any-source| {host-source*<source-host-ip>*}} {{*<destination>*<*destination-wildcard>*}|any-destination| {host-destination *<destination-host-ip>*}} [*<icmp-type>* [*<icmp-code>*]] [precedence *<precedence>*][tos*<tos>*][time-range*<time-range-name>*]	Creates an extended name-based MAC-ICMP access rule; the no form command deletes this name-based extended MAC-ICMP access rule.
[no]{deny|permit}{any-source-mac|{host-source-mac *<host_smac>*}|{*<smac>*<*smac-mask>*}} {any-destination-mac|{host-destination-mac *<host_dmac>*}|{*<dmac>*<*dmac-mask>*}}igmp {{*<source>*<*source-wildcard>*}|any-source| {host-source*<source-host-ip>*}} {{*<destination>*<*destination-wildcard>*}|any-destination| {host-destination *<destination-host-ip>*}} [*<igmp-type>*] [precedence *<precedence>*] [tos *<tos>*][time-range*<time-range-name>*]	Creates an extended name-based MAC-IGMP access rule; the no form command deletes this name-based extended MAC-IGMP access rule.
[no]{deny|permit}{any-source-mac|{host-source-mac<host_smac>}|{<smac><smac-mask>}} {any-destination-mac|{host-destination-mac <host_dmac>}|{<dmac><dmac-mask>}}tcp {{<source><source-wildcard>}|any-source| {host-source<source-host-ip>}} [s-port {*<port1>* | range *<sPortMin>* *<sPortMax>*}] {{<destination><destination-wildcard>}|any-destination| {host-destination <destination-host-ip>}} [d-port {*<port3>* | range *<dPortMin>* *<dPortMax>*}] [ack+fin+psh+rst+urg+syn] [precedence<precedence>][tos<tos>][time-range*<time-r*	Creates an extended name-based MAC-TCP access rule; the no form command deletes this name-based extended MAC-TCP access rule.

***ange-name>*]**	
[no]{deny|permit}{any-source-mac|{host-source-mac<*host_smac*>}|{<*smac*><*smac-mask*>}} {any-destination-mac|{host-destination-mac <*host_dmac*>}|{<*dmac*><*dmac-mask*>}}udp {{<*source*><*source-wildcard*>}|any-source| {host-source<*source-host-ip*>}} [s-port {<*port1*> | range <*sPortMin*> <*sPortMax*>}] {{<*destination*><*destination-wildcard*>}|any-destination| {host-destination <*destination-host-ip*>}} [d-port {<*port3*> | range <*dPortMin*> <*dPortMax*>}] [precedence <*precedence*>] [tos <*tos*>][time-range<*time-range-name*>]	Creates an extended name-based MAC-UDP access rule; the no form command deletes this name-based extended MAC-UDP access rule.
[no]{deny|permit}{any-source-mac|{host-source-mac<*host_smac*>}|{<*smac*><*smac-mask*>}} {any-destination-mac|{host-destination-mac <*host_dmac*>}|{<*dmac*><*dmac-mask*>}} {eigrp|gre|igrp|ip|ipinip|ospf|{<*protocol-num*>}} {{<*source*><*source-wildcard*>}|any-source| {host-source<*source-host-ip*>}} {{<*destination*><*destination-wildcard*>}|any-destination| {host-destination<*destination-host-ip*>}} [precedence<*precedence*>][tos<*tos*>][time-range<*time-range-name*>]	Creates an extended name-based access rule for the other IP protocol; the no form command deletes this name-based extended access rule.

c. Exit MAC-IP Configuration Mode

Command	Explanation
Extended name-based MAC-IP access Mode	
exit	Quit extended name-based MAC-IP access mode.

（10） Configuring a numbered standard IPv6 access-list

Command	Explanation
Global Mode	
ipv6 access-list <*num*> {deny | permit} {{<*sIPv6Addr*> <*sPrefixlen*>} | any-source | {host-source <*sIpv6Addr*>}} **no ipv6 access-list <*num*>**	Creates a numbered standard IPv6 access-list, if the access-list already exists, then a rule will add to the current access-list; the "**no access-list <*num*>**" command deletes a numbered standard IPv6 access-list.

（11） Configuring a numbered extensive IPv6 access-list

Command	Explanation
Global Mode	
ipv6 access-list <num-ext> {deny | permit} icmp {{<sIPv6Prefix/sPrefixlen>} | any-source | {host-source <sIPv6Addr>}} {<dIPv6Prefix/dPrefixlen> | any-destination | {host-destination <dIPv6Addr>}} [<icmp-type> [<icmp-code>]] [dscp <dscp>] [flow-label <fl>][time-range<time-range-name>] **ipv6 access-list <num-ext> {deny | permit} tcp {{<sIPv6Prefix/<sPrefixlen>} | any-source | {host-source <sIPv6Addr>}} [s-port {<sPort> | range <sPortMin> <sPortMax>}] {{< dIPv6Prefix/<dPrefixlen>} | any-destination | {host-destination <dIPv6Addr>}} [dPort {<dPort> | range <dPortMin> <dPortMax>}] [syn | ack | urg | rst | fin | psh] [dscp <dscp>] [flow-label <flowlabel>]**[**time-range<*time-range-name*>**]** **ipv6 access-list <num-ext> {deny | permit} udp {{<sIPv6Prefix/<sPrefixlen>} | any-source | {host-source <sIPv6Addr>}} [s-port {<sPort> | range <sPortMin> <sPortMax>}] {{<dIPv6Prefix/<dPrefixlen>} | any-destination | {host-destination <dIPv6Addr>}} [dPort {<dPort> | range <dPortMin> <dPortMax>}] [dscp <dscp>] [flow-label <flowlabel>][time-range<time-range-name>]** **ipv6 access-list <num-ext> {deny | permit} <next-header> {<sIPv6Prefix/sPrefixlen> | any-source | {host-source <sIPv6Addr>}} {<dIPv6Prefix/dPrefixlen> | any-destination | {host-destination <dIPv6Addr>}} [dscp <dscp>] [flow-label <fl>]**[**time-range<*time-range-name*>**]** **no ipv6 access-list <num>**	Creates a numbered extended IPv6 access-list, if the access-list already exists, then a rule will add to the current access-list; the no command deletes a numbered standard IPv6 access-list.

（12）Configuring a standard IPv6 access-list based on nomenclature

a. Create a standard IPv6 access-list based on nomenclature

Command	Explanation
Global Mode	
ipv6 access-list standard <name> **no ipv6 access-list standard <name>**	Creates a standard IP access-list based on nomenclature; the no command delete the name-based standard IPv6 access-list.

b. Specify multiple permit or deny rules

Command	Explanation
Standard IPv6 ACL Mode	

[no] {deny | permit} {{<sIPv6Prefix/sPrefixlen>} | any-source | {host-source <sIPv6Addr> }}	Creates a standard name-based IPv6 access rule; the no form command deletes the name-based standard IPv6 access rule.

c. Exit name-based standard IP ACL configuration mode

Command	Explanation
Standard IPv6 ACL Mode	
exit	Exits name-based standard IPv6 ACL configuration mode.

（13）Configuring an name-based extended IPv6 access-list

a. Create an extended IPv6 access-list basing on nomenclature

Command	Explanation
Global Mode	
ipv6 access-list extended <name> **no ipv6 access-list extended <name>**	Creates an extended IPv6 access-list basing on nomenclature; the no command deletes the name-based extended IPv6 access-list.

b. Specify multiple permit or deny rules

Command	Explanation
Extended IPv6 ACL Mode	
[no] {deny | permit} icmp {{<sIPv6Prefix/sPrefixlen>} | any-source | {host-source <sIPv6Addr>}} {<dIPv6Prefix/dPrefixlen> | any-destination | {host-destination <dIPv6Addr>}} [<icmp-type> [<icmp-code>]] [dscp <dscp>] [flow-label <flowlabel>] [time-range<time-range-name>]	Creates an extended name-based ICMP IPv6 access rule; the no form command deletes this name-based extended IPv6 access rule.
[no] {deny | permit} tcp {<sIPv6Prefix/sPrefixlen> | any-source | {host-source <sIPv6Addr>}} [s-port {<sPort> | range <sPortMin> <sPortMax>}] {<dIPv6Prefix/dPrefixlen> | any-destination | {host-destination <dIPv6Addr>}} [d-port {<dPort> | range <dPortMin> <dPortMax>}] [syn | ack | urg | rst | fin | psh] [dscp <dscp>] [flow-label <fl>] [**time-range<*time-range-name*>**]	Creates an extended name-based TCP IPv6 access rule; the no form command deletes this name-based extended IPv6 access rule.
[no] {deny | permit} udp	Creates an extended name-based UDP IPv6 access

{<sIPv6Prefix/sPrefixlen> | any-source | {host-source <sIPv6Addr>}} [s-port {<sPort> | range <sPortMin> <sPortMax>}] {<dIPv6Prefix/dPrefixlen> | any-destination | {host-destination <dIPv6Addr>}} [d-port {<dPort> | range <dPortMin> <dPortMax>}] [dscp <dscp>] [flow-label <fl>][time-range<*time-range-name*>]	rule; the no form command deletes this name-based extended IPv6 access rule.
[no] {deny | permit} <proto> {<sIPv6Prefix/sPrefixlen> | any-source | {host-source <sIPv6Addr>}} {<dIPv6Prefix/dPrefixlen> | any-destination | {host-destination <dIPv6Addr>}} [dscp <dscp>] [flow-label <flowlabel>] [time-range<time-range-name>]	Creates an extended name-based IPv6 access rule for other IPv6 protocols; the no form command deletes this name-based extended IPv6 access rule.
[no] {deny | permit} {<sIPv6Prefix/sPrefixlen> | any-source | {host-source <sIPv6Addr>}} {<dIPv6Prefix/dPrefixlen> | any-destination | {host-destination <dIPv6Addr>}} [dscp <dscp>] [flow-label <flowlabel>] [time-range<time-range-name>]	Creates an extended name-based IPv6 access rule; the no form command deletes this name-based extended IPv6 access rule.

c. Exit extended IPv6 ACL configuration mode

Command	Explanation
Extended IPv6 ACL Mode	
exit	Exits extended name-based IPv6 ACL configuration mode.

2. Configuring packet filtering function

(1) Enable global packet filtering function

Command	Explanation
Global Mode	
firewall enable	Enables global packet filtering function.
firewall disable	Disables global packet filtering function.

(2) Configure ACL deny preemption function globally

Command	Explanation
Global Mode	
[no] access-list deny-preemption	Enable deny-preemption function, the no command disables deny-preemption function.

3. Configuring time range function

（1）Create the name of the time range

Command	Explanation
Global Mode	
time-range *<time_range_name>*	Create a time range named time_range_name.
no time-range *<time_range_name>*	Stop the time range function named time_range_name.

（2）Configure periodic time range

Command	Explanation
Time range Mode	
absolute-periodic {Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday} *<start_time>* to {Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday} *<end_time>*	Configure the time range for the request of the week, and every week will run by the time range.
periodic {{Monday+Tuesday+Wednesday+Thursday+Friday+Saturday+Sunday} | daily | weekdays | weekend} *<start_time>* to *<end_time>*	
[no] absolute-periodic {Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday} *<start_time>* to {Monday | Tuesday | Wednesday | Thursday | Friday | Saturday | Sunday} *<end_time>*	Stop the function of the time range in the week.
[no] periodic {{Monday+Tuesday+Wednesday+Thursday+Friday+Saturday+Sunday} | daily | weekdays | weekend} *<start_time>* to *<end_time>*	

（3）Configure absolute time range

Command	Explanation
Global Mode	
absolute start *<start_time> <start_data>* [end *<end_time> <end_data>*]	Configure absolute time range.
[no] absolute start *<start_time> <start_data>* [end *<end_time> <end_data>*]	Stop the function of the time range.

4. Bind access-list to a specific direction of the specified port.

Command	Explanation
Physical Port Mode/VLAN Interface Mode	

{ip|ipv6|mac|mac-ip} access-group *<acl-name>* {in|out} [traffic-statistic] **no {ip|ipv6|mac|mac-ip} access-group *<acl-name>* {in|out}**	Apply an access-list to the ingress or egress direction on the port; the no command deletes the access-list bound to the port.

5. Show configuration and statistic information of ACL

Command	Explanation
Admin Mode	
show access-group in (interface {Ethernet | Ethernet IFNAME})	Show the ACL binding status on the port.

6. Clear the filtering information of the specified port

Command	Explanation
Admin Mode	
clear access-group (in | out) statistic interface { *<interface-name>* | ethernet *<interface-name>* }	Clear the filtering information of the specified port.

6.1.3 ACL Example

Scenario 1:

The user has the following configuration requirement: port 10 of the switch connects to 10.0.0.0/24 segment, ftp is not desired for the user.

Configuration description:

1. Create a proper ACL
2. Configuring packet filtering function
3. Bind the ACL to the port

The configuration steps are listed below:

Switch(config)#access-list 110 deny tcp 10.0.0.0 0.0.0.255 any-destination d-port 21
Switch(config)#firewall enable
Switch(config)#interface ethernet 1/0/10
Switch(Config-If-Ethernet1/0/10)#ip access-group 110 in
Switch(Config-If-Ethernet1/0/10)#exit
Switch(config)#exit

Configuration result:

Switch#show firewall
Firewall status: enable.
Switch#show access-lists
access-list 110(used 1 time(s)) 1 rule(s)
access-list 110 deny tcp 10.0.0.0 0.0.0.255 any-destination d-port 21

Switch#show access-group interface ethernet 1/0/10
interface name:Ethernet1/0/10
the ingress acl use in firewall is 110, traffic-statistics Disable.

Scenario 2:

The configuration requirement is stated as below: The switch should drop all the 802.3 datagram with 00-12-11-23-xx-xx as the source MAC address coming from interface 10.

Configuration description:

1. Create the corresponding MAC ACL.
2. Configure datagram filtering.
3. Bind the ACL to the related interface.

The configuration steps are listed as below.

Switch(config)#access-list 1100 deny 00-12-11-23-00-00 00-00-00-00-ff-ff any-destination-mac untagged-802-3
Switch(config)#access-list 1100 deny 00-12-11-23-00-00 00-00-00-00-ff-ff any tagged-802
Switch(config)#firewall enable
Switch(config)#interface ethernet1/0/10
Switch(Config-If-Ethernet1/0/10)#mac access-group 1100 in
Switch(Config-If-Ethernet1/0/10)#exit
Switch(config)#exit

Configuration result:

Switch#show firewall
Firewall Status: Enable.

Switch #show access-lists
access-list 1100(used 1 time(s))
access-list 1100 deny 00-12-11-23-00-00 00-00-00-00-ff-ff
any-destination-mac
untagged-802-3
access-list 1100 deny 00-12-11-23-00-00 00-00-00-00-ff-ff
any-destination-mac
Switch #show access-group interface ethernet 1/0/10
interface name:Ethernet1/0/10
MAC Ingress access-list used is 1100,traffic-statistics Disable.

Scenario 3:

The configuration requirement is stated as below: The MAC address range of the network connected to the interface 10 of the switch is 00-12-11-23-xx-xx, and IP network is 10.0.0.0/24. FTP should be disabled and ping requests from outside network should be disabled.

Configuration description:

1. Create the corresponding access list.
2. Configure datagram filtering.

3. Bind the ACL to the related interface.

The configuration steps are listed as below.

Switch(config)#access-list 3110 deny 00-12-11-23-00-00 00-00-00-00-ff-ff any-destination-mac tcp 10.0.0.0 0.0.0.255 any-destination d-port 21

Switch(config)#access-list 3110 deny any-source-mac 00-12-11-23-00-00 00-00-00-00-ff-ff icmp any-source 10.0.0.0 0.0.0.255

Switch(config)#firewall enable

Switch(config)#interface ethernet 1/0/10

Switch(Config-If-Ethernet1/0/10)#mac-ip access-group 3110 in

Switch(Config-Ethernet1/0/10)#exit

Switch(config)#exit

Configuration result:

Switch#show firewall

Firewall Status: Enable.

Switch#show access-lists

access-list 3110(used 1 time(s))

access-list 3110 deny 00-12-11-23-00-00 00-00-00-00-ff-ff

any-destination-mac

tcp 10.0.0.0 0.0.0.255 any-destination d-port 21

access-list 3110 deny any-source-mac 00-12-11-23-00-00 00-00-00-00-ff-ff icmp any-source 10.0.0.0 0.0.0.255

Switch #show access-group interface ethernet 1/0/10

interface name:Ethernet1/0/10

MAC-IP Ingress access-list used is 3110, traffic-statistics Disable.

Scenario 4:

The configuration requirement is stated as below: IPv6 protocol runs on the interface 600 of the switch. And the IPv6 network address is 2003:1:1:1::0/64. Users in the 2003:1:1:1:66::0/80 subnet should be disabled from accessing the outside network.

Configuration description:

1. Create the corresponding access list.
2. Configure datagram filting.
3. Bind the ACL to the related interface.

The configuration steps are listed as below.

Switch(config)#ipv6 access-list 600 permit 2003:1:1:1:66::0/80 any-destination

Switch(config)#ipv6 access-list 600 deny 2003:1:1:1::0/64 any-destination

Switch(config)#firewall enable

Switch(config)#interface ethernet 1/0/10

Switch(Config-If-Ethernet1/0/10)#ipv6 access-group 600 in

Switch(Config-If-Ethernet1/0/10)#exit
Switch(config)#exit
Configuration result:
Switch#show firewall
Firewall Status: Enable.

Switch#show ipv6 access-lists
Ipv6 access-list 600(used 1 time(s))
ipv6 access-list 600 deny 2003:1:1:1::0/64 any-source
ipv6 access-list 600 permit 2003:1:1:1:66::0/80 any-source

Switch #show access-group interface ethernet 1/0/10
interface name:Ethernet1/0/10
IPv6 Ingress access-list used is 600, traffic-statistics Disable.

Scenario 5:

The configuration requirement is stated as below: The interface 1, 2, 5, 7 belongs to vlan100, Hosts with 192.168.0.1 as its IP address should be disabled from accessing the listed interfaces.

Configuration description:

1. Create the corresponding access list.
2. Configure datagram filtering.
3. Bind the ACL to the related interface.

The configuration steps are listed as below.
Switch (config)#firewall enable
Switch (config)#vlan 100
Switch (Config-Vlan100)#switchport interface ethernet 1/0/1;2;5;7
Switch (Config-Vlan100)#exit
Switch (config)#access-list 1 deny host-source 192.168.0.1
Switch (config)#interface ethernet1/0/1;2;5;7
Switch (config-if-port-range)#ip access-group 1 in
Switch (Config-if-Vlan100)#exit

Configuration result:
Switch (config)#show access-group interface vlan 100
Interface VLAN 100:
Ethernet1/0/1: IP Ingress access-list used is 1, traffic-statistics Disable.
Ethernet1/0/2: IP Ingress access-list used is 1, traffic-statistics Disable.
Ethernet1/0/5: IP Ingress access-list used is 1, traffic-statistics Disable.
Ethernet1/0/7: IP Ingress access-list used is 1, traffic-statistics Disable.

6.1.4 ACL Troubleshooting

Checking for entries in the ACL is done in a top-down order and ends whenever an entry is matched.

- Default rule will be used only if no ACL is bound to the incoming direction of the port, or no ACL entry is matched.Each ingress port can bind one MAC-IP ACL, one IP ACL, one MAC ACL, one IPv6 standard ACL (via the physical interface mode or Vlan interface mode).
- When binding four ACL and packet matching several ACL at the same time, the priority relations are as follows in a top-down order. If the priority is same, then the priority of configuration at first is higher.
  - Ingress IPv6 ACL
  - Ingress MAC-IP ACL
  - Ingress IP ACL
  - Ingress MAC ACL
- The number of ACLs that can be successfully bound depends on the content of the ACL bound and the hardware resource limit. Users will be prompted if an ACL cannot be bound due to hardware resource limitation.
  - If an access-list contains same filtering information but conflicting action rules, binding to the port will fail with an error message. For instance, configuring 'permit tcp any any-destination' and 'deny tcp any any-destination' at the same time is not permitted.
- Viruses such as 'worm.blaster' can be blocked by configuring ACL to block specific ICMP packets or specific TCP or UDP port packet.
- If the physical mode of an interface is TRUNK, ACL can only be configured through physical interface mode.
- ACL configured in the physical mode can only be disabled in the physical mode. Those configured in the VLAN interface configuration mode can only be disabled in the VLAN interface mode.
- When a physical interface is added into or removed from a VLAN (with the trunk interfaces as exceptions), ACL configured in the corresponding VLAN will be bound or unbound respectively. If ACL configured in the target VLAN, which is configured in VLAN interface mode, conflicts with existing ACL configuration on the interface, which is configured in physical interface mode, the configuration will fail to effect.
- When no physical interfaces are configured in the VLAN, the ACL configuration of the VLAN will be removed. And it can not recover if new interfaces are added to the VLAN.
- When the interface mode is changed from access mode to trunk mode, the ACL configured in VLAN interface mode which is bound to physical interface will be removed. And when the interface mode is changed from trunk mode to access mode, ACL configured in VLAN1 interface mode will be bound to the physical interface. If binding fails, the changing will fail either.

6.2 Self-defined ACL

6.2.1 Introduction to Self-defined ACL

ACL (Access Control Lists) is a packet filtering mechanism implemented by switch, providing network access control by granting or denying access the switches, effectively safeguarding the security of networks. The user can set a set of rules according to some information specific to packets, each rule describes the action for a packet with certain information matched: 'permit' or 'deny'. The user can apply such rules to the incoming direction of switch ports, so that data streams of specified ports must comply with the ACL rules assigned.

Self-defined ACL means that users can configure several self-defined windows as the matching field when users configure ACL. Self-defined windows do not specify which field definitely, but specify the offset in a packet and ignore the meaning of field. It matches the data at offset position which begins to fix the byte length according to the value and mask configuration.

6.2.1.1 Standard Self-defined ACL Template

Standard self-defined ACL can configure 16 windows. Each window can specify offset, its value from 0 to 31, unit is 2Bytes, namely, 0 means 0Bytes offset and 1 means 2Bytes offset. Besides, offset is according to the start offset position.

A standard self-defined ACL template should be configured for the offset configuration of every window before configuring the standard self-defined ACL list. This template is global and takes effect to all standard self-defined ACL list. Standard self-defined ACL template can configure the offset for 16 windows at most. The window which is not configured is not available, that means it cannot transmit configuration successfully if the standard self-defined ACL use this window. When a window in the template is configured, it cannot be modified if the standard self-defined ACL rule is configured with this window. But the standard self-defined ACL rule is not configured; the window can be reconfigured, modified or deleted. When uses it with other functions (such as AM, ARP-GUARP, anti-arpscan), please reduce using the window and keep them for 8 or less than it.

6.2.1.2 Digital Self-defined ACL

Digital self-defined ACL can configure multi-ACL lists and each of them can configure multi-rules.

The number of the issued lists is according to the type of the card. One rule can configure value and mask for 16 windows at most. The length of every window is 2Bytes; the name range of the self-defined ACL list is <1200-1299>.

6.2.1.3 Named Self-defined ACL

Named self-defined ACL can configure multi-ACL lists and each of them can configure multi-rules. The number of the issued lists is according to the type of the card. One rule can configure value and mask for 16 windows at most. The length of every window is 2Bytes; the length of the name string is 1-64, and the string cannot include numbers only.

6.2.1.4 Self-defined ACL Configuration Transmitting

6.2.1.5 Special Explanation

6.2.2 Self-defined ACL Configuration

Task list of self-defined ACL configuration:

1. Configure the offset template
2. Configure the digital standard userdefined acl
3. Configure the named standard userdefined acl
4. Bind the userdefined acl rule to the port

1. Configure the offset template

Command	Explanation
Global Mode	
userdefined-access-list standard offset [window1 <*offset*>] [window2 <*offset*>] [window3 <*offset*>] [window4 <*offset*>] [window5 <*offset*>] [window6 <*offset*>] [window7 <*offset*>] [window8 <*offset*>] [window9 <*offset*>] [window10 <*offset*>] [window11 <*offset*>] [window12 <*offset*>] [window13 <*offset*>] [window14 <*offset*>] [window15 <*offset*>] [window16 <*offset*>] **no userdefined-access-list standard offset [window1*]* [window2] [window3] [window4] [window5] *[*window6] [window7] [window8] [window9] [window10*]* [window11]**	Create a standard self-defined ACL template. If the template exists, the corresponding window of the template can be modified; the no command deletes the window of the standard self-defined ACL template. If the window is not specified, the standard self-defined ACL template will be deleted.

[window12] [window13] [window14] *[*window15] [window16]	

2. Configure the digital standard userdefined acl

Command	Explanation
Global Mode	
userdefined-access-list standard <num> {deny | permit} [packet-type ipv4 | ipv6 | l2-eth2 | l2-llc | l2-snap | mpls] *[*window1 <value> <*mask*>] [window2 <*value*> <mask>] *[*window3 <value> <*mask*>] [window4 <*value*> <mask>] *[*window5 <value> <*mask*>] [window6 <*value*> <mask>] *[*window7 <value> <*mask*>] [window8 <*value*> <mask>] [window9 <value> <mask>] [window10 <value> <mask>] *[*window11 <value> <mask>] [window12 <value> <mask>] [window13 <value> <mask>] *[*window14 <value> <mask>] [window15 <value> <mask>] [window16 <*value*> <*mask*>] **no userdefined-access-list <num>**	Create a numbered standard self-defined ACL. If the standard self-defined ACL exists, then a rule will be added to the ACL. The no command deletes a numbered standard self-defined ACL.

3. Configure the named standard userdefined acl

Command	Explanation
Global Mode	
[no] udf-access-list standard <name>	Create a userdefined access-list. The no command deletes it.
Name Userdefined Access-list Mode.	
**[no] {deny | permit*}* [packet-type ipv4 | ipv6 | l2-eth2 | l2-llc | l2-snap | mpls] [window1 <*value*> <mask>] *[*window2 <value> <*mask*>] [window3 <*value*> <mask>] *[*window4 <value> <*mask*>] [window5 <*value*> <mask>] *[*window6 <value> <*mask*>] [window7 <*value*> <mask>] *[*window8 <value> <*mask*>] [window9 <value> <mask>*]* [window10 <value> <mask>] [window11 <*value*> <mask>] [window12 <value> <mask>*]* [window13 <value> <mask>] [window14 <*value*> <mask>] [window15 <value> <mask>*]* [window16 <*value*> <*mask*>]	Add a rule table entry in the userdefined acl list. The no command deletes the access-list.

4. Bind the userdefined acl rule to the port

Command	Explanation

Port Mode	
[no] userdefined access-group {<*name*>|<num>} {in} [traffic-statistic]	Apply userdefined-access-list to one direction of the port. Decide whether the statistical counter should be added to the ACL according to the options. The no command deletes the configuration bound to the port.

6.2.3 Self-defined ACL Example

Scenario 1:

The user has the following configuration requirement: The packet whose first and second byte is 0x0003 is not allowed forwarding on port 1 of the switch.

Configuration description**:**

Create a self-defined ACL template according to condition

1. Create a corresponding self-defined ACL
2. Bind the self-defined ACL to the port
3. The configuration steps are listed **below:**
4. Switch(config)#userdefined-access-list standard offset window1 0
5. Switch(config)#userdefined-access-list standard 1200 deny window1 0003 FFFF

Switch(config)#

Switch(config)#firewall enable

Switch(config)#interface ethernet 1/0/1

Switch(config-if-ethernet1/0/1)#userdefined access-group 1200 in

Switch(config)#exit

Configuration result:

Switch #show **access-lists**

userdefined-access-list standard 1200(used 1 time(s)) 1 rule(s)

rule ID 1: deny window1 0003 ffff

Switch#show access-group interface ethernet 1/0/1

interface name:Ethernet1/0/1

Userdefined Ingress access-list used is 1200, traffic-statistics Disable.

6.2.4 Self-defined ACL Troubleshooting

☞ The troubleshooting of self-defined ACL can be dealt with according to the ordinary ACL. If the operation is wrong, the system will show the detailed error information.

6.3 802.1x

6.3.1 Introduction to 802.1x

The 802.1x protocol originates from 802.11 protocol, the wireless LAN protocol of IEEE, which is designed to provide a solution to doing authentication when users access a wireless LAN. The LAN defined in IEEE 802 LAN protocol does not provide access authentication, which means as long as the users can access a LAN controlling device (such as a LAN Switch), they will be able to get all the devices or resources in the LAN. There was no looming danger in the environment of LAN in those primary enterprise networks.

However, along with the boom of applications like mobile office and service operating networks, the service providers should control and configure the access from user. The prevailing application of WLAN and LAN access in telecommunication networks, in particular, make it necessary to control ports in order to implement the user-level access control. And as a result, IEEE LAN/WAN committee defined a standard, which is 802.1x, to do Port-Based Network Access Control. This standard has been widely used in wireless LAN and ethernet.

"Port-Based Network Access Control" means to authenticate and control the user devices on the level of ports of LAN access devices. Only when the user devices connected to the ports pass the authentication, can they access the resources in the LAN, otherwise, the resources in the LAN won't be available.

6.3.1.1 The Authentication Structure of 802.1x

The system using 802.1x has a typical Client/Server structure, which contains three entities (as illustrated in the next figure): Supplicant system, Authenticator system, and Authentication server system.

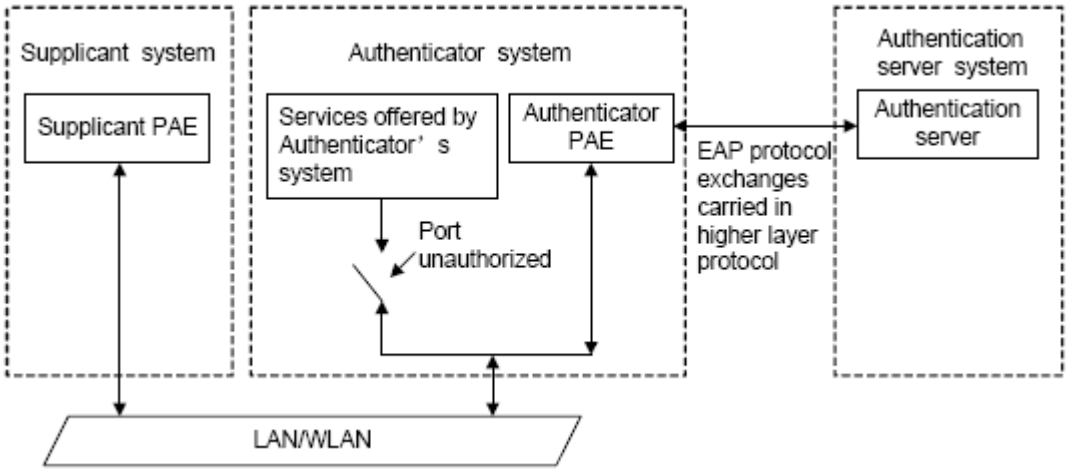


Figure 6-1 The Authentication Structure of 802.1x

☞ The supplicant system is an entity on one end of the LAN segment, should be authenticated by the access controlling unit on the other end of the link. A Supplicant

system usually is a user terminal device. Users start 802.1x authentication by starting supplicant system software. A supplicant system should support EAPOL (Extensible Authentication Protocol over LAN).

- ☞ The authenticator system is another entity on one end of the LAN segment to authenticate the supplicant systems connected. An authenticator system usually is a network device supporting 802,1x protocol, providing ports to access the LAN for supplicant systems. The ports provided can either be physical or logical.
- ■ The authentication server system is an entity to provide authentication service for authenticator systems. The authentication server system is used to authenticate and authorize users, as well as does fee-counting, and usually is a RADIUS (Remote Authentication Dial-In User Service) server, which can store the relative user information, including username, password and other parameters such as the VLAN and ports which the user belongs to.

The three entities above concerns the following basic concepts: PAE of the port, the controlled ports and the controlled direction.

1. PAE

PAE (Port Access Entity) is the entity to implement the operation of algorithms and protocols.

- ☞ The PAE of the supplicant system is supposed to respond the authentication request from the authenticator systems and submit user's authentication information to the authenticator system. It can also send authentication request and off-line request to authenticator.
- ☞ The PAE of the authenticator system authenticates the supplicant systems needing to access the LAN via the authentication server system, and deal with the authenticated/unauthenticated state of the controlled port according to the result of the authentication. The authenticated state means the user is allowed to access the network resources, the unauthenticated state means only the EAPOL messages are allowed to be received and sent while the user is forbidden to access network resources.

2. controlled/uncontrolled ports

The authenticator system provides ports to access the LAN for the supplicant systems. These ports can be divided into two kinds of logical ports: controlled ports and uncontrolled ports.

- ☞ The uncontrolled port is always in bi-directionally connected status, and mainly used to transmit EAPOL protocol frames, to guarantee that the supplicant systems can always send or receive authentication messages.
- ☞ The controlled port is in connected status authenticated to transmit service messages. When unauthenticated, no message from supplicant systems is allowed to be received.
- ☞ The controlled and uncontrolled ports are two parts of one port, which means each frame

reaching this port is visible on both the controlled and uncontrolled ports.

3. Controlled direction

In unauthenticated status, controlled ports can be set as unidirectional controlled or bi-directionally controlled.

- ☞ When the port is bi-directionally controlled, the sending and receiving of all frames is forbidden.
- ☞ When the port is unidirectional controlled, no frames can be received from the supplicant systems while sending frames to the supplicant systems is allowed.

Notes: At present, this kind of switch only supports unidirectional control.

6.3.1.2 The Work Mechanism of 802.1x

IEEE 802.1x authentication system uses EAP (Extensible Authentication Protocol) to implement exchange of authentication information between the supplicant system, authenticator system and authentication server system.

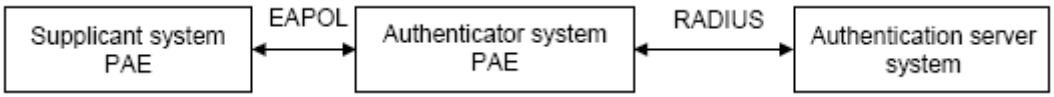


Figure 6-2 the Work Mechanism of 802.1x

- ☞ EAP messages adopt EAPOL encapsulation format between the PAE of the supplicant system and the PAE of the authenticator system in the environment of LAN.
- ☞ Between the PAE of the authenticator system and the RADIUS server, there are two methods to exchange information: one method is that EAP messages adopt EAPOR (EAP over RADIUS) encapsulation format in RADIUS protocol; the other is that EAP messages terminate with the PAE of the authenticator system, and adopt the messages containing RAP (Password Authentication Protocol) or CHAP (Challenge Handshake Authentication Protocol) attributes to do the authentication interaction with the RADIUS server.
- ☞ When the user pass the authentication, the authentication server system will send the relative information of the user to authenticator system, the PAE of the authenticator system will decide the authenticated/unauthenticated status of the controlled port according to the authentication result of the RADIUS server.

6.3.1.3 The Encapsulation of EAPOL Messages

1. The Format of EAPOL Data Packets

EAPOL is a kind of message encapsulation format defined in 802.1x protocol, and is mainly used to transmit EAP messages between the supplicant system and the authenticator system in order to allow the transmission of EAP messages through the LAN. In IEEE 802/Ethernet LAN environment, the format of EAPOL packet is illustrated in the next figure. The beginning of the

EAPOL packet is the Type/Length domain in MAC frames.

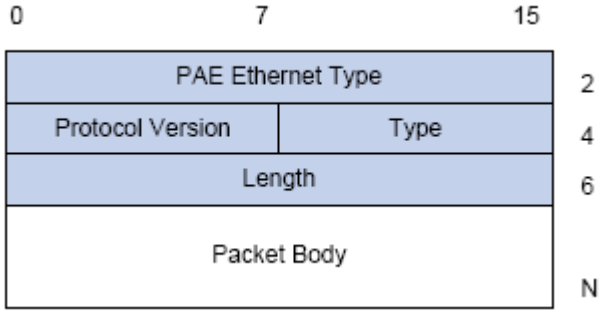


Figure 6-3 the Format of EAPOL Data Packet

PAE Ethernet Type: Represents the type of the protocol whose value is 0x888E.

Protocol Version: Represents the version of the protocol supported by the sender of EAPOL data packets.

Type: represents the type of the EAPOL data packets, including:

- EAP-Packet (whose value is 0x00): the authentication information frame, used to carry EAP messages. This kind of frame can pass through the authenticator system to transmit EAP messages between the supplicant system and the authentication server system.
- EAPOL-Start (whose value is 0x01): the frame to start authentication.
- EAPOL-Logoff (whose value is 0x02): the frame requesting to quit.
- EAPOL-Key (whose value is 0x03): the key information frame.
- EAPOL-Encapsulated-ASF-Alert (whose value is 0x04): used to support the Alerting messages of ASF (Alert Standard Forum). This kind of frame is used to encapsulate the relative information of network management such as all kinds of alerting information, terminated by terminal devices.

Length: represents the length of the data, that is, the length of the “Packet Body”, in byte. There will be no following data domain when its value is 0.

Packet Body: represents the content of the data, which will be in different formats according to different types.

2. The Format of EAP Data Packets

When the value of Type domain in EAPOL packet is EAP-Packet, the Packet Body is in EAP format (illustrated in the next figure).

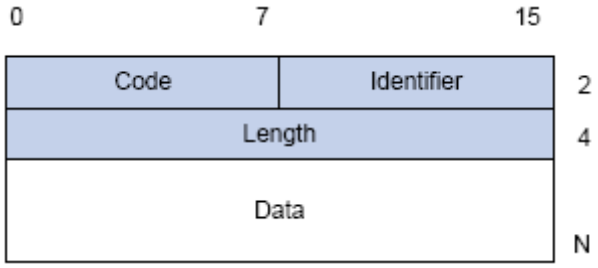


Figure 6-4 the Format of EAP Data Packets

Code: specifies the type of the EAP packet. There are four of them in total: Request（1）,Response（2）,Success（3）,Failure（4）.

- There is no Data domain in the packets of which the type is Success or Failure, and the value of the Length domains in such packets is 4.
- The format of Data domains in the packets of which the type is Request and Response is illustrated in the next figure. Type is the authentication type of EAP, the content of Type data depends on the type. For example, when the value of the type is 1, it means Identity, and is used to query the identity of the other side. When the type is 4, it means MD5-Challenge, like PPP CHAP protocol, contains query messages.

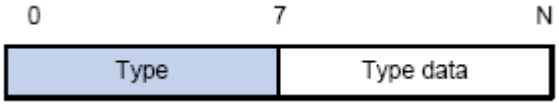


Figure 6-5 the Format of Data Domain in Request and Response Packets

Identifier: to assist matching the Request and Response messages.

Length: the length of the EAP packet, covering the domains of Code, Identifier, Length and Data, in byte.

Data: the content of the EAP packet, depending on the Code type.

6.3.1.4 The Encapsulation of EAP Attributes

RADIUS adds two attribute to support EAP authentication: EAP-Message and Message-Authenticator. Please refer to the Introduction of RADIUS protocol in "AAA-RADIUS-HWTACACS operation" to check the format of RADIUS messages.

1. EAP-Message

As illustrated in the next figure, this attribute is used to encapsulate EAP packet, the type code is 79, String domain should be no longer than 253 bytes. If the data length in an EAP packet is larger than 253 bytes, the packet can be divided into fragments, which then will be encapsulated in several EAP-Messages attributes in their original order.

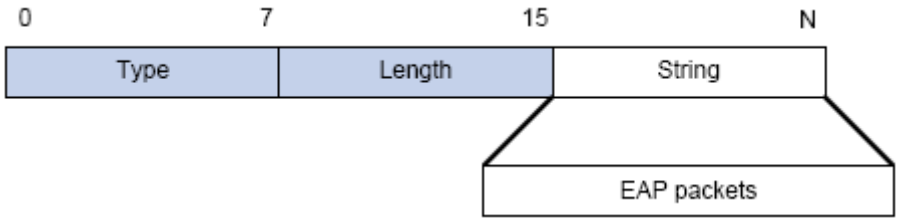


Figure 6-6 the Encapsulation of EAP-Message Attribute

2. Message-Authenticator

As illustrated in the next figure, this attribute is used in the process of using authentication methods like EAP and CHAP to prevent the access request packets from being eavesdropped. Message-Authenticator should be included in the packets containing the EAP-Message attribute, or the packet will be dropped as an invalid one.

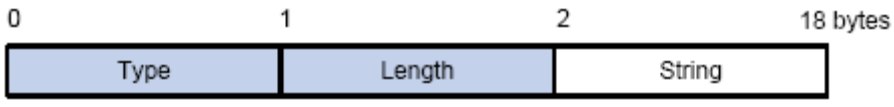


Figure 6-7 Message-Authenticator Attribute

6.3.1.5 The Authentication Methods of 802.1x

The authentication can either be started by supplicant system initiatively or by devices. When the device detects unauthenticated users to access the network, it will send supplicant system EAP-Request/Identity messages to start authentication. On the other hand, the supplicant system can send EAPOL-Start message to the device via supplicant software.

802.1 x systems supports EAP relay method and EAP termination method to implement authentication with the remote RADIUS server. The following is the description of the process of these two authentication methods, both started by the supplicant system.

6.3.1.5.1 EAP Relay Mode

EAP relay is specified in IEEE 802.1x standard to carry EAP in other high-level protocols, such as EAP over RADIUS, making sure that extended authentication protocol messages can reach the authentication server through complicated networks. In general, EAP relay requires the RADIUS server to support EAP attributes: EAP-Message and Message-Authenticator.

EAP is a widely-used authentication frame to transmit the actual authentication protocol rather than a special authentication mechanism. EAP provides some common function and allows the authentication mechanisms expected in the negotiation, which are called EAP Method. The advantage of EAP lies in that EAP mechanism working as a base needs no adjustment when a new authentication protocol appears. The following figure illustrates the protocol stack of EAP authentication method.

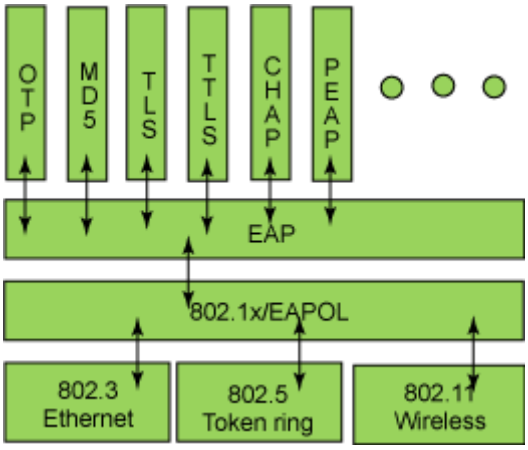


Figure 6-8 the Protocol Stack of EAP Authentication Method

By now, there are more than 50 EAP authentication methods has been developed, the differences among which are those in the authentication mechanism and the management of keys. The 4 most common EAP authentication methods are listed as follows:

- **EAP-MD5**
- **EAP-TLS**（Transport Layer Security）
- **EAP-TTLS**（Tunneled Transport Layer Security）
- **PEAP**（Protected Extensible Authentication Protocol）

They will be described in detail in the following part.

Attention:

- ☞ The switch, as the access controlling unit of Pass-through, will not check the content of a particular EAP method, so can support all the EAP methods above and all the EAP authentication methods that may be extended in the future.
- ☞ In EAP relay, if any authentication method in EAP-MD5, EAP-TLS, EAP-TTLS and PEAP is adopted, the authentication methods of the supplicant system and the RADIUS server should be the same.

1. EAP-MD5 Authentication Method

EAP-MD5 is an IETF open standard which providing the least security, since MD5 Hash function is vulnerable to dictionary attacks.

The following figure illustrated the basic operation flow of the EAP-MD5 authentication method.

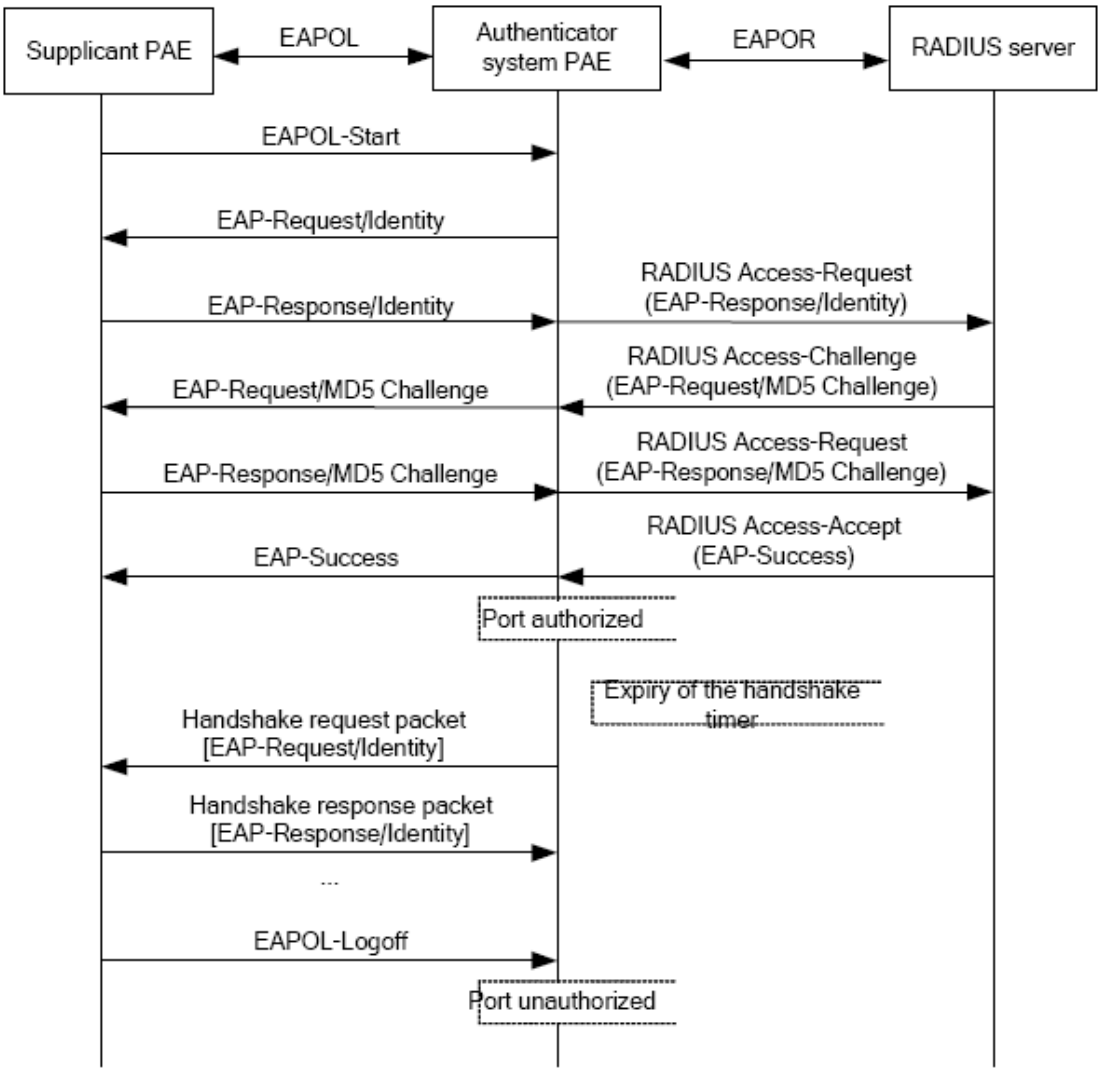


Figure 6-9 the Authentication Flow of 802.1x EAP-MD5

2. EAP-TLS Authentication Method

EAP-TLS is brought up by Microsoft based on EAP and TLS protocols. It uses PKI to protect the id authentication between the supplicant system and the RADIUS server and the dynamically

generated session keys, requiring both the supplicant system and the Radius authentication server to possess digital certificate to implement bidirectional authentication. It is the earliest EAP authentication method used in wireless LAN. Since every user should have a digital certificate, this method is rarely used practically considering the difficult maintenance. However it is still one of the safest EAP standards, and enjoys prevailing supports from the vendors of wireless LAN hardware and software.

The following figure illustrates the basic operation flow of the EAP-TLS authentication method.

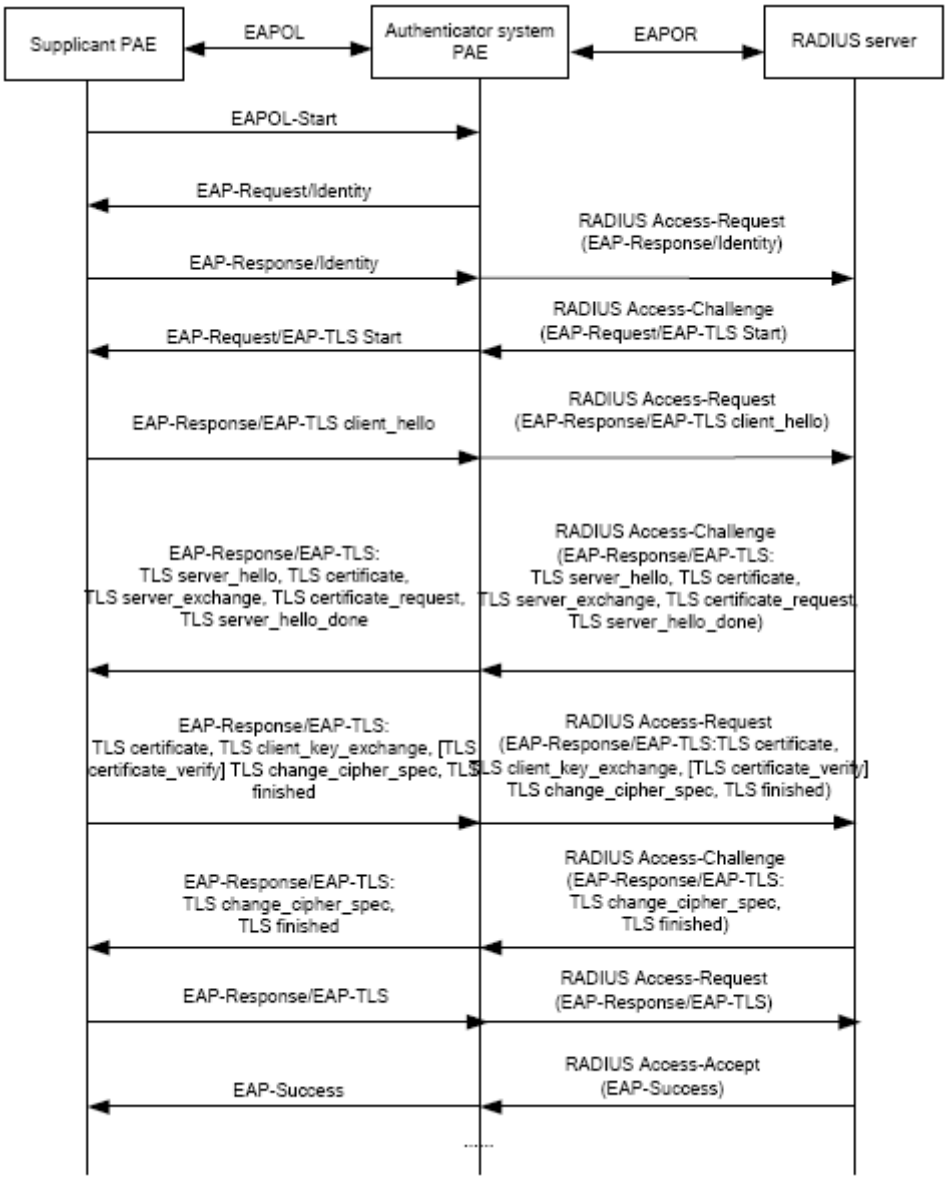


Figure 6-10 the Authentication Flow of 802.1x EAP-TLS

3. EAP-TTLS Authentication Method

EAP-TTLS is a product of the cooperation of Funk Software and Certicom. It can provide an authentication as strong as that provided by EAP-TLS, but without requiring users to have their own digital certificate. The only request is that the Radius server should have a digital certificate. The authentication of users' identity is implemented with passwords transmitted in a safely

encrypted tunnel established via the certificate of the authentication server. Any kind of authentication request including EAP, PAP and MS-CHAPV2 can be transmitted within TTLS tunnels.

4. PEAP Authentication Method

EAP-PEAP is brought up by Cisco, Microsoft and RAS Security as a recommended open standard. It has long been utilized in products and provides very good security. Its design of protocol and security is similar to that of EAP-TTLS, using a server's PKI certificate to establish a safe TLS tunnel in order to protect user authentication.

The following figure illustrates the basic operation flow of PEAP authentication method.

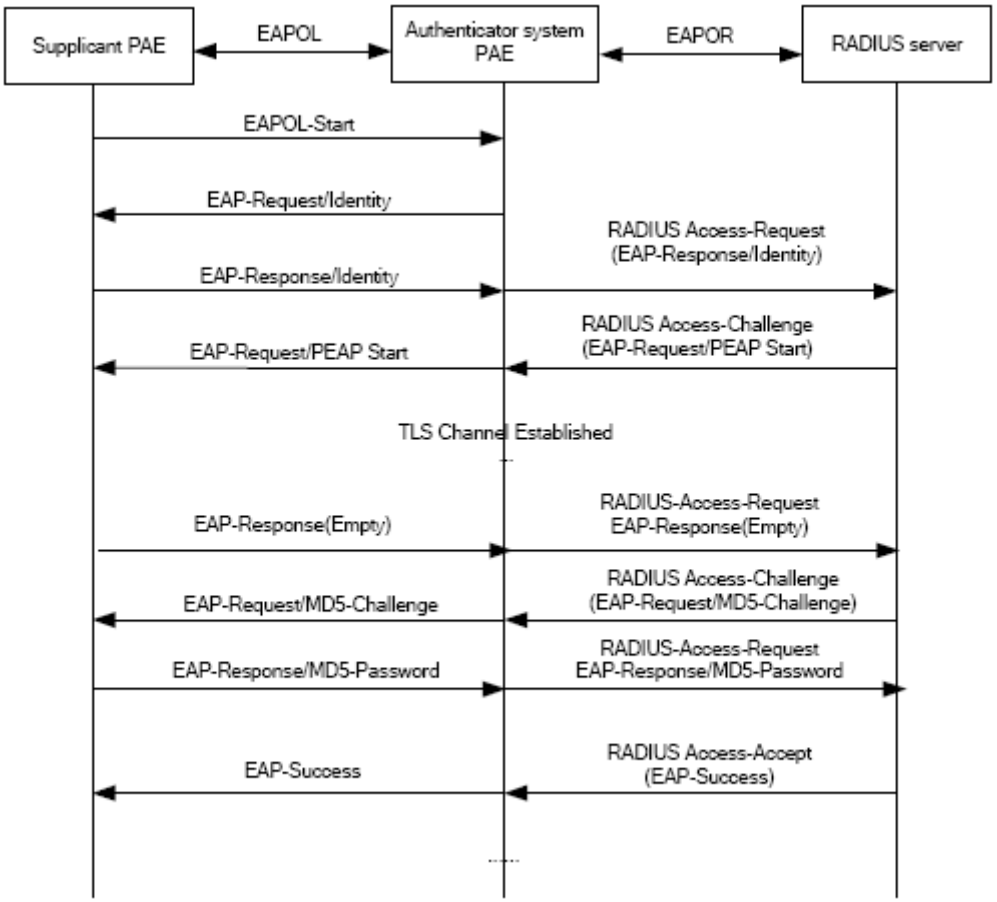


Figure 6-11 the Authentication Flow of 802.1x PEAP

6.3.1.5.2 EAP Termination Mode

In this mode, EAP messages will be terminated in the access control unit and mapped into RADIUS messages, which is used to implement the authentication, authorization and fee-counting. The basic operation flow is illustrated in the next figure.

In EAP termination mode, the access control unit and the RADIUS server can use PAP or CHAP authentication method. The following figure will demonstrate the basic operation flow using CHAP authentication method.

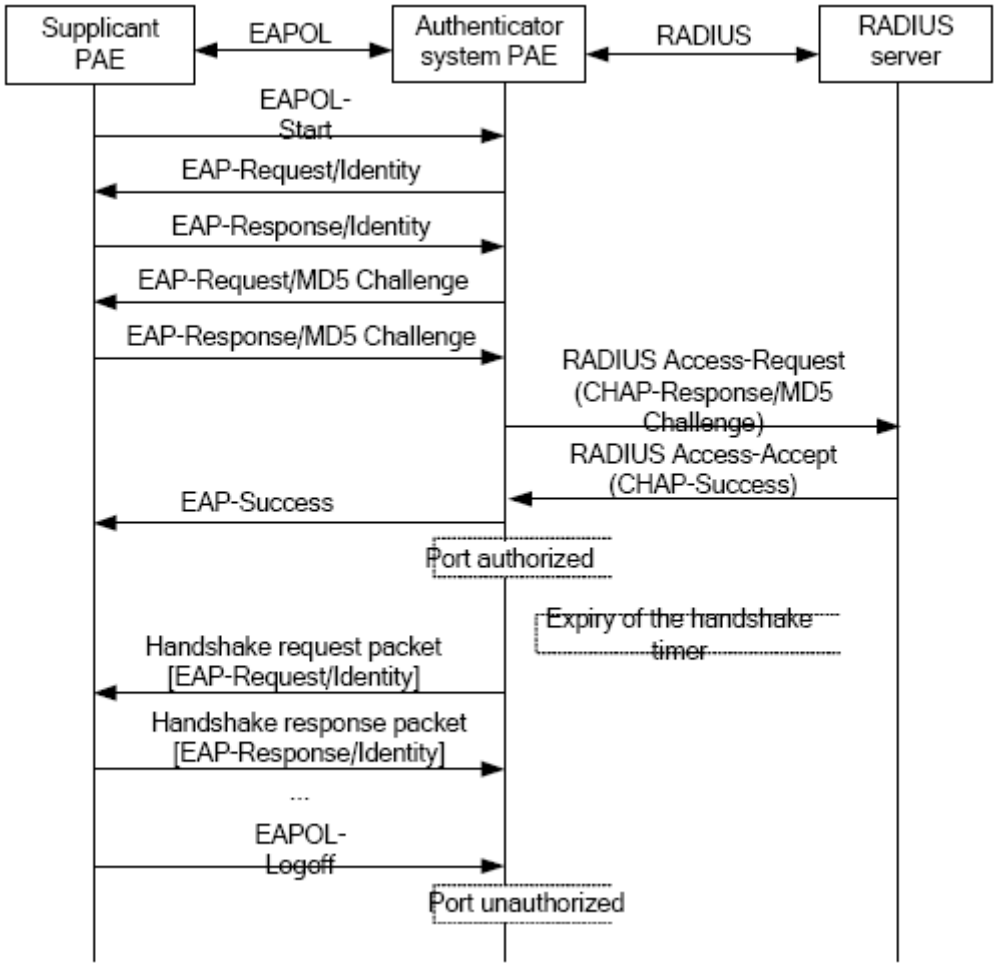


Figure 6-12 the Authentication Flow of 802.1x EAP Termination Mode

6.3.1.6 The Extension and Optimization of 802.1x

Besides supporting the port- based access authentication method specified by the protocol, devices also extend and optimize it when implementing the EAP relay mode and EAP termination mode of 802.1x.

- ☞ Supports some applications in the case of which one physical port can have more than one users
- ☞ There are three access control methods (the methods to authenticate users): port-based, MAC-based and user-based (IP address+ MAC address+ port).
  - When the port-based method is used, as long as the first user of this port passes the authentication, all the other users can access the network resources without being authenticated. However, once the first user is offline, the network won't be available to all the other users.
  - When the MAC-based method is used, all the users accessing a port should be authenticated separately, only those pass the authentication can access the network,

while the others can not. When one user becomes offline, the other users will not be affected.

- When the user-based (IP address+ MAC address+ port) method is used, all users can access limited resources before being authenticated. There are two kinds of control in this method: standard control and advanced control. The user-based standard control will not restrict the access to limited resources, which means all users of this port can access limited resources before being authenticated. The user-based advanced control will restrict the access to limited resources, only some particular users of the port can access limited resources before being authenticated. Once those users pass the authentication, they can access all resources.

Attention: when using private supplicant systems, user-based advanced control is recommended to effectively prevent ARP cheat.

The maximum number of the authenticated users can be 4000, but less than 2000 will be preferred

6.3.1.7 The Features of VLAN Allocation

1. Auto VLAN

Auto VLAN feature enables RADIUS server to change the VLAN to which the access port belongs, based on the user information and the user access device information. When an 802.1x user passes authentication on the server, the RADIUS server will send the authorization information to the device, if the RADIUS server has enabled the VLAN-assigning function, then the following attributes should be included in the Access-Accept messages:

- ☞ Tunnel-Type = VLAN (13)
- ☞ Tunnel-Medium-Type = 802 (6)
- ☞ Tunnel-Private-Group-ID = VLANID

The VLANID here means the VID of VLAN, ranging from 1 to 4094. For example, Tunnel-Private-Group-ID = 30 means VLAN 30.

When the switch receives the assigned Auto VLAN information, the current Access port will leave the VLAN set by the user and join Auto VLAN.

Auto VLAN won't change or affect the port's configuration. But the priority of Auto VLAN is higher than that of the user-set VLAN, that is Auto VLAN is the one takes effect when the authentication is finished, while the user-set VLAN do not work until the user become offline.

Notes: At present, Auto VLAN can only be used in the port-based access control mode, and on the ports whose link type is Access.

2. Guest VLAN

Guest VLAN feature is used to allow the unauthenticated user to access some specified resources.

The user authentication port belongs to a default VLAN (Guest VLAN) before passing the

802.1x authentication, with the right to access the resources within this VLAN without authentication. But the resources in other networks are beyond reach. Once authenticated, the port will leave Guest VLAN, and the user can access the resources of other networks.

In Guest VLAN, users can get 802.1x supplicant system software, update supplicant system or update some other applications (such as anti-virus software, the patches of operating system). The access device will add the port into Guest VLAN if there is no supplicant getting authenticated successfully in a certain stretch of time because of lacking exclusive authentication supplicant system or the version of the supplicant system being too low.

Once the 802.1x feature is enabled and the Guest VLAN is configured properly, a port will be added into Guest VLAN, just like Auto VLAN, if there is no response message from the supplicant system after the device sends more authentication-triggering messages than the upper limit (EAP-Request/Identity) from the port.

- ☞ The authentication server assigns an Auto VLAN, and then the port leaves Guest VLAN and joins the assigned Auto VLAN. When the user becomes offline, the port will be allocated to the specified Guest VLAN again.
- ☞ The authentication server assigns an Auto VLAN, and then the port leaves Guest VLAN and joins the specified VLAN. When the user becomes offline, the port will be allocated to the specified Guest VLAN again.

6.3.2 802.1x Configuration Task List

802.1x Configuration Task List:

1. Enable IEEE 802.1x function
2. Access management unit property configuration

1) Configure access management method for the port: MAC-based only

2) Configure expanded 802.1x function

3) Configure DHCP passthrough function of the port

3. User access devices related property configuration (optional)

1. Enable 802.1x function

Command	Explanation
Global Mode	
dot1x enable [vlan-pool] **no dot1x enable [vlan-pool]**	Enables the 802.1x function in the switch and ports; the no command disables the 802.1x function.
free-resource destination [ipv4|ipv6] <*prefix*> /<*mask*> **no free-resource destination [ipv4|ipv6] <*prefix*> /<*mask*>**	Sets free access network resource for unauthorized dot1x user. The no command close the resource.
dot1x unicast enable ***no dot1x unicast enable***	Enable the 802.1x unicast passthrough function of switch; the no operation of this command will disable this function.

2. Access management unit property configuration

1) Configure port access management method

Command	Explanation

Port Mode	
dot1x max-user macbased <*number*> **no dot1x max-user macbased**	Sets the maximum number of access users for the specified port; the no command restores the default setting of allowing 50000 user.
dot1x guest-vlan <vlanID> **no dot1x guest-vlan**	Configure the guest vlan of the appointed port. The no command deletes it.

2) Configure expanded 802.1x function

Command	Explanation
Global Mode	
dot1x macbased port-down-flush **no dot1x macbased port-down-flush**	Enables this command, when the dot1x certification according to mac is down, delete the user who passed the certification of the port; The no command does not make the down operation.
dot1x eapor enable **no dot1x eapor enable**	Enables the EAP relay authentication function in the switch; the no command sets EAP local end authentication.

3) Configure DHCP passthrough function of the port

Command	Explanation
Port Mode	
dot1x dhcp passthrough **no dot1x dhcp passthrough**	Enables DHCP passthrough function of global mode on a switch; the no operation of this command will disable the function.

3. Supplicant related property configuration

Command	Explanation
Global Mode	
dot1x max-req <count> no dot1x max-req	Sets the number of EAP request/MD5 frame to be sent before the switch re-initials authentication on no supplicant response, the no command restores the default setting.
dot1x re-authentication no dot1x re-authentication	Enables periodical supplicant authentication; the no command disables this function.
dot1x timeout quiet-period <seconds> no dot1x timeout quiet-period	Sets time to keep silent on port authentication failure; the no command restores the default value.
dot1x timeout re-authperiod <seconds> no dot1x timeout re-authperiod	Sets the supplicant re-authentication interval; the no command restores the default setting.
dot1x timeout tx-period <seconds> no dot1x timeout tx-period	Sets the interval for the supplicant to re-transmit EAP request/identity frame; the no command restores the default setting.
dot1x re-authenticate [interface <interface-name>]	Enables IEEE 802.1x re-authentication (no wait timeout requires) for all ports or a specified port.

6.3.3 802.1x Application Example

6.3.3.1 Examples of Guest Vlan Applications

6.3.3.2 Examples of IPv4 Radius Applications

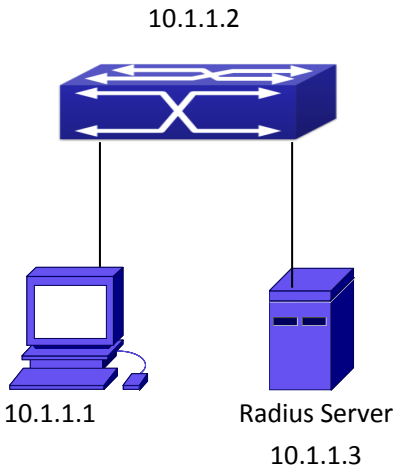


Figure 6-13 IEEE 802.1x Configuration Example Topology

The PC is connecting to port 1/0/2 of the switch; IEEE 802.1x authentication is enabled on port1/0/2; the access mode is the default MAC-based authentication. The switch IP address is 10.1.1.2. Any port other than port 1/0/2 is used to connect to RADIUS authentication server, which has an IP address of 10.1.1.3, and use the default port 1812 for authentication and port 1813 for accounting. IEEE 802.1x authentication client software is installed on the PC and is used in IEEE 802.1x authentication.

The configuration procedures are listed below:

```
Switch(config)#interface vlan 1
Switch(Config-if-vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-vlan1)#exit
Switch(config)#radius-server authentication host 10.1.1.3
Switch(config)#radius-server accounting host 10.1.1.3
Switch(config)#radius-server key test
Switch(config)#aaa enable
Switch(config)#aaa-accounting enable
Switch(config)#dot1x enable
Switch(config)#interface ethernet 1/0/2
Switch(Config-Ethernet1/0/2)#dot1x enable
Switch(Config-Ethernet1/0/2)#exit
```

6.3.3.3 Examples of IPv6 Radius Application

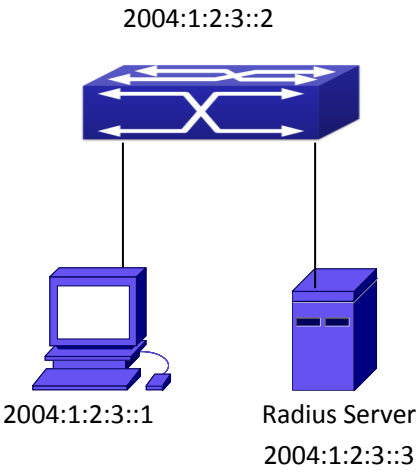


Figure 6-14 IPv6 Radius

Connect the computer to the interface 1/0/2 of the switch, and enable IEEE802.1x on interface1/0/2. Use MAC based authentication. Configure the IP address of the switch as 2004:1:2:3::2, and connect the switch with any interface except interface 1/0/2 to the RADIUS authentication server. Configure the IP address of the RADIUS server to be 2004:1:2:3::3. Use the default ports 1812 and 1813 for authentication and accounting respectively. Install the IEEE802.1x authentication client software on the computer, and use the client for IEEE802.1x authentication.

The detailed configurations are listed as below:

Switch(config)#interface vlan 1
Switch(Config-if-vlan1)#ipv6 address 2004:1:2:3::2/64
Switch(Config-if-vlan1)#exit
Switch(config)#radius-server authentication host 2004:1:2:3::3
Switch(config)#radius-server accounting host 2004:1:2:3::3
Switch(config)#radius-server key test
Switch(config)#aaa enable
Switch(config)#aaa-accounting enable
Switch(config)#dot1x enable
Switch(config)#interface ethernet 1/0/2
Switch(Config-If-Ethernet1/0/2)#dot1x enable
Switch(Config-If-Ethernet1/0/2)#exit

6.3.4 802.1x Troubleshooting

It is possible that 802.1x be configured on ports and 802.1x authentication be set to auto, t switch can’t be to authenticated state after the user runs 802.1x supplicant software. Here are some possible causes and solutions:

- ☞ If 802.1x cannot be enabled for a port, make sure the port is not executing MAC binding, or configured as a port aggregation. To enable the 802.1x authentication, the above functions must be disabled.
- ☞ If the switch is configured properly but still cannot pass through authentication, connectivity between the switch and RADIUS server, the switch and 802.1x client should be verified, and the port and VLAN configuration for the switch should be checked, too.
- ☞ Check the event log in the RADIUS server for possible causes. In the event log, not only unsuccessful logins are recorded, but prompts for the causes of unsuccessful login. If the event log indicates wrong authenticator password, radius-server key parameter shall be modified; if the event log indicates no such authenticator, the authenticator needs to be added to the RADIUS server; if the event log indicates no such login user, the user login ID and password may be wrong and should be verified and input again.

6.4 The Number Limitation Function of MAC and IP in Port, VLAN

6.4.1 Introduction to the Number Limitation Function of MAC and IP in Port, VLAN

MAC address list is used to identify the mapping relationship between the destination MAC addresses and the ports of switch. There are two kinds of MAC addresses in the list: static MAC address and dynamic MAC address. The static MAC address is set by users, having the highest priority (will not be overwritten by dynamic MAC address), and will always be effective; dynamic MAC address is learnt by the switch through transmitting data frames, and will only be effective in a specific time range. When the switch receives a data framed waiting to be transmitted, it will study the source MAC address of the data frame, build a mapping relationship with the receiving port, and then look up the MAC address list for the destination MAC address. If any matching list entry is found, the switch will transmit the data frame via the corresponding port, or, the switch will broadcast the data frame over the VLAN it belongs to. If the dynamically learnt MAC address matches no transmitted data in a long time, the switch will delete it from the MAC address list.

Usually the switch supports both the static configuration and dynamic study of MAC address, which means each port can have more than one static set MAC addresses and dynamically learnt MAC addresses, and thus can implement the transmission of data traffic between port and known MAC addresses. When a MAC address becomes out of date, it will be dealt with broadcast. No number limitation is put on MAC address of the ports of our current switches; every port can have several MAC addressed either by configuration or study, until the hardware list entries are exhausted. To avoid too many MAC addresses of a port, we should limit the number of MAC

addresses a port can have.

For each INTERFACE VLAN, there is no number limitation of IP; the upper limit of the number of IP is the upper limit of the number of user on an interface, which is, at the same time, the upper limit of ARP and ND list entry. There is no relative configuration command can be used to control the sent number of these list entries. To enhance the security and the controllability of our products, we need to control the number of MAC address on each port and the number of ARP, ND on each INTERFACE VLAN. The number of static or dynamic MAC address on a port should not exceed the configuration. The number of user on each VLAN should not exceed the configuration, either.

Limiting the number of MAC and ARP list entry can avoid DOS attack to a certain extent. When malicious users frequently do MAC or ARP cheating, it will be easy for them to fill the MAC and ARP list entries of the switch, causing successful DOS attacks.

To summer up, it is very meaningful to develop the number limitation function of MAC and IP in port, VLAN. Switch can control the number of MAC address of ports and the number ARP, ND list entry of ports and VLAN through configuration commands.

Limiting the number of dynamic MAC and IP of ports:

1. Limiting the number of dynamic MAC. If the number of dynamically learnt MAC address by the switch is already larger than or equal with the max number of dynamic MAC address, then shutdown the MAC study function on this port, otherwise, the port can continue its study.

2. Limiting the number of dynamic IP. If the number of dynamically learnt ARP and ND by the switch is already larger than or equal with the max number of dynamic ARP and ND, then shutdown the ARP and ND study function of this port, otherwise, the port can continue its study.

Limiting the number of MAC, ARP and ND of interfaces:

1. Limiting the number of dynamic MAC. If the number of dynamically learnt MAC address by the VLAN of the switch is already larger than or equal with the max number of dynamic MAC address, then shutdown the MAC study function of all the ports in this VLAN, otherwise, all the ports in this VLAN can continue their study (except special ports).

2. Limiting the number of dynamic IP. If the number of dynamically learnt ARP and ND by the switch is already larger than or equal with the max number of dynamic ARP and ND, then the VLAN will not study any new ARP or ND, otherwise, the study can be continued.

6.4.2 The Number Limitation Function of MAC and IP in Port, VLAN Configuration Task Sequence

1. Enable the number limitation function of MAC and IP on ports
2. Enable the number limitation function of MAC and IP in VLAN
3. Configure the timeout value of querying dynamic MAC
4. Configure the violation mode of ports
5. Display and debug the relative information of number limitation of MAC and IP on ports

1. Enable the number limitation function of MAC and IP on ports

Command	Explanation
Port configuration mode	

switchport mac-address dynamic maximum *<value>* **no switchport mac-address dynamic maximum**	Enable and disable the number limitation function of MAC on the ports. **Notice:** This command is not supported by the switch.
switchport arp dynamic maximum *<value>* **no switchport arp dynamic maximum**	Enable and disable the number limitation function of ARP on the ports.
switchport nd dynamic maximum *<value>* **no switchport nd dynamic maximum**	Enable and disable the number limitation function of ND on the ports.

2. Enable the number limitation function of MAC and IP in VLAN

Command	Explanation
VLAN configuration mode	
vlan mac-address dynamic maximum *<value>* **no vlan mac-address dynamic maximum**	Enable and disable the number limitation function of MAC in the VLAN. **Notice:** This command is not supported by the switch.
Interface configuration mode	
ip arp dynamic maximum *<value>* **no ip arp dynamic maximum**	Enable and disable the number limitation function of ARP in the VLAN.
ipv6 nd dynamic maximum *<value>* **no ipv6 nd dynamic maximum**	Enable and disable the number limitation function of NEIGHBOR in the VLAN.

3. Configure the timeout value of querying dynamic MAC

Command	Explanation
Global configuration mode	
mac-address query timeout <seconds>	Configure the timeout value of querying dynamic MAC. **Notice:** This command is not supported by the switch.

4. Configure the violation mode of ports

Command	Explanation
Port mode	
switchport mac-address violation {protect / shutdown} [recovery <5-3600>] **no switchport mac-address violation**	Set the violation mode of the port, the no command restores the violation mode to **protect**.

5. Display and debug the relative information of number limitation of MAC and IP on ports

Command	Explanation
Admin mode	

show mac-address dynamic count {vlan <*vlan-id*> | interface ethernet <portName> }	Display the number of dynamic MAC in corresponding ports and VLAN. **Notice:** This command is not supported by the switch.
show arp-dynamic count **{vlan <vlan-id> | interface ethernet <*portName*> }**	Display the number of dynamic ARP in corresponding ports and VLAN.
show nd-dynamic count {vlan <*vlan-id*> | interface ethernet <*portName*> }	Display the number of dynamic NEIGHBOUR in corresponding ports and VLAN.
debug switchport mac count **no debug switchport mac count**	All kinds of debug information when limiting the number of MAC on ports. **Notice:** This command is not supported by the switch.
debug switchport arp count **no debug switchport arp count**	All kinds of debug information when limiting the number of ARP on ports.
debug switchport nd count **no debug switchport nd count**	All kinds of debug information when limiting the number of NEIGHBOUR on ports.
debug vlan mac count **no debug vlan mac count**	All kinds of debug information when limiting the number of MAC in VLAN. **Notice:** This command is not supported by the switch.
debug ip arp count **no debug ip arp count**	All kinds of debug information when limiting the number of ARP in VLAN.
debug ipv6 nd count **no debug ipv6 nd count**	All kinds of debug information when limiting the number of MACin VLAN.

6.4.3 The Number Limitation Function of MAC and IP in Port, VLAN Typical Examples

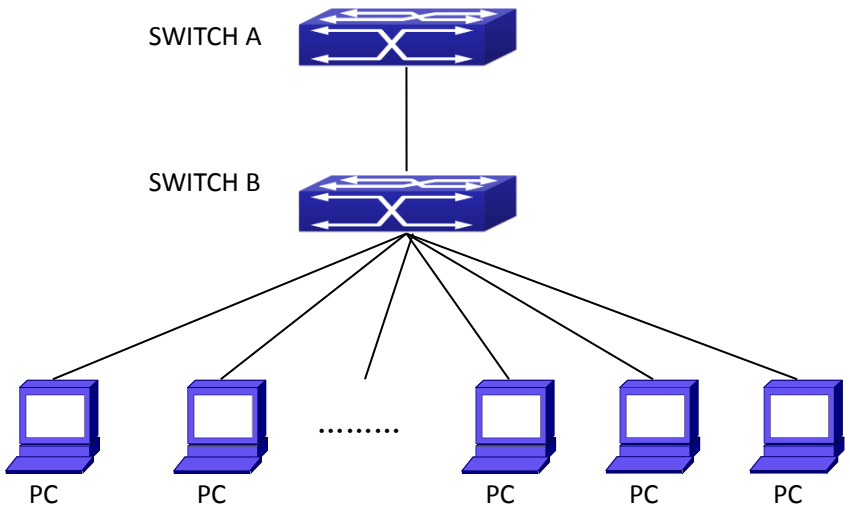


Figure 6-195 The Number Limitation of MAC and IP in Port, VLAN Typical Configuration Example

In the network topology above, SWITCH B connects to many PC users, before enabling the number limitation function of MAC and IP in Port, VLAN, if the system hardware has no other limitation, SWTICH A and SWTICH B can get the MAC, ARP, ND list entries of all the PC, so limiting the MAC, ARP list entry can avoid DOS attack to a certain extent. When malicious users frequently do MAC, ARP cheating, it will be easy for them to fill the MAC, ARP list entries of the switch, causing successful DOS attacks. Limiting the MAC, ARP, ND list entry can prevent DOS attack.

On port 1/0/1 of SWITCH A, set the max number can be learnt of dynamic MAC address as 20, dynamic ARP address as 20, NEIGHBOR list entry as 10. In VLAN 1, set the max number of dynamic MAC address as 30, of dynamic ARP address as 30, NEIGHBOR list entry as 20.

SWITCH A configuration task sequence:
Switch (config)#interface ethernet 1/0/1
Switch (Config-If-Ethernet1/0/1)#switchport mac-address dynamic maximum 20
Switch (Config-If-Ethernet1/0/1)#switchport arp dynamic maximum 20
Switch (Config-If-Ethernet1/0/1)#switchport nd dynamic maximum 10
Switch (Config-if-Vlan1)#vlan mac-address dynamic maximum 30

6.4.4 The Number Limitation Function of MAC and IP in Port, VLAN Troubleshooting Help

The number limitation function of MAC and IP in Port, VLAN is disabled by default, if users need to limit the number of user accessing the network, they can enable it. If the number limitation function of MAC address can not be configured, please check whether Spanning-tree, dot1x, TRUNK is running on the switch and whether the port is configured as a MAC-binding port. The number limitation function of MAC address is mutually exclusive to these configurations, so if the users need to enable the number limitation function of MAC address on the port, they should check these functions mentioned above on this port are disabled.

If all the configurations are normal, after enabling the number limitation function of MAC and IP in Port, VLAN, users can use debug commands to debug every limitation, check the details of number limitations and judge whether the number limitation function is correct. If there is any problem, please sent result to technical service center.

6.5 AM

6.5.1 Introduction to AM Function

AM (Access Management) means that when a switch receives an IP or ARP message, it will compare the information extracted from the message (such as source IP address or source MAC-IP address) with the configured hardware address pool. If there is an entry in the address pool matching the information (source IP address or source MAC-IP address), the message will be forwarded, otherwise, dumped. The reason why source-IP-based AM should be supplemented by source-MAC-IP-based AM is that IP address of a host might change. Only with a bound IP, can users change the IP of the host into forwarding IP, and hence enable the messages from the host to be forwarded by the switch. Given the fact that MAC-IP can be exclusively bound with a host, it is necessary to make MAC-IP bound with a host for the purpose of preventing users from maliciously modifying host IP to forward the messages from their hosts via the switch.

With the interface-bound attribute of AM, network mangers can bind the IP (MAC-IP) address of a legal user to a specified interface. After that, only the messages sending by users with specified IP (MAC-IP) addresses can be forwarded via the interface, and thus strengthen the monitoring of the network security.

6.5.2 AM Function Configuration Task List

1. Enable AM function
2. Enable AM function on an interface
3. Configure the forwarding IP
4. Configure the forwarding MAC-IP
5. Delete all of the configured IP or MAC-IP or both
6. Display relative configuration information of AM

1. Enable AM function

Command	Explanation
Global Mode	
am enable **no am enable**	Globally enable or disable AM function.

2. Enable AM function on an interface

Command	Explanation
Port Mode	
am port **no am port**	Enable/disable AM function on the port. When the AM function is enabled on the port, no IP or ARP message will be forwarded by default.

3. Configure the forwarding IP

Command	Explanation

Port Mode	
am ip-pool *<ip-address> <num>* **no am ip-pool *<ip-address> <num>***	Configure the forwarding IP of the port.

4. Configure the forwarding MAC-IP

Command	Explanation
Port Mode	
am mac-ip-pool *<mac-address> <ip-address>* **no am mac-ip-pool *<mac-address> <ip-address>***	Configure the forwarding MAC-IP of the port.

5. Delete all of the configured IP or MAC-IP or both

Command	Explanation
Global Mode	
no am all [ip-pool|mac-ip-pool]	Delete MAC-IP address pool or IP address pool or both pools configured by all users.

6. Display relative configuration information of AM

Command	Explanation
Global Configuration Mode	
show am [interface *<interface-name>*]	Display the AM configuration information of one port or all ports.

6.5.3 AM Function Example

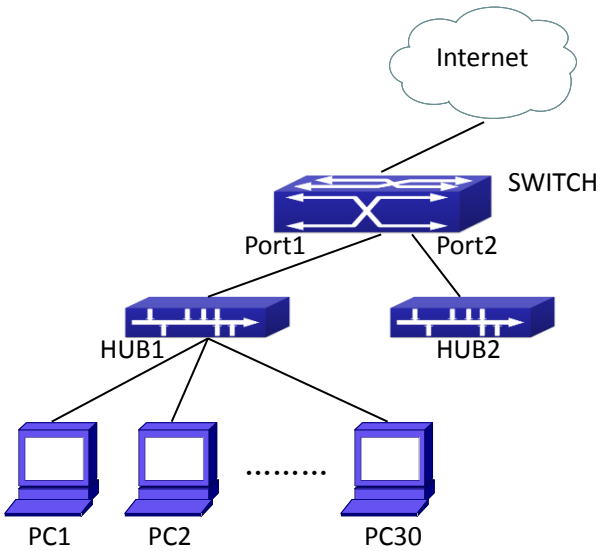


Figure 6-16 a typical configuration example of AM function

In the topology above, 30 PCs, after converged by HUB1, connect with interface1 on the switch. The IP addresses of these 30 PCs range from 100.10.10.1 to 100.10.10.30. Considering

security, the system manager will only take user with an IP address within that range as legal ones. And the switch will only forward data packets from legal users while dumping packets from other users.

According to the requirements mentioned above, the switch can be configured as follows:

```
Switch(config)#am enable
Switch(config)#interface ethernet1/0/1
Switch(Config-If-Ethernet 1/0/1)#am port
Switch(Config-If-Ethernet 1/0/1)#am ip-pool 10.10.10.1 10
```

6.5.4 AM Function Troubleshooting

AM function is disabled by default, and after it is enabled, relative configuration of AM can be made.

Users can view the current AM configuration with “show am” command, such as whether the AM is enabled or not, and AM information on each interface, they can also use “**show am [interface *<interface-name>*]**” command to check the AM configuration information on a specific interface.

If any operational error happens, the system will display detailed corresponding prompt.

6.6 Security Feature

6.6.1 Introduction to Security Feature

Before introducing the security features, we here first introduce the DoS. The DoS is short for Denial of Service, which is a simple but effective destructive attack on the internet. The server under DoS attack will drop normal user data packet due to non-stop processing the attacker’s data packet, leading to the denial of the service and worse can lead to leak of sensitive data of the server.

Security feature refers to applications such as protocol check which is for protecting the server from attacks such as DoS. The protocol check allows the user to drop matched packets based on specified conditions. The security features provide several simple and effective protections against Dos attacks while acting no influence on the linear forwarding performance of the switch.

6.6.2 Security Feature Configuration

6.6.2.1 Prevent IP Spoofing Function Configuration Task Sequence

1. Enable the IP spoofing function.

Command		Explanation
Global Mode		
[no] dosattack-check srcip-equal-dstip enable		Enable/disable th source address is address.

6.6.2.2 Prevent TCP Unauthorized Label Attack Function Configuration Task Sequence

1. Enable the anti TCP unauthorized label attack function
2. Enable Checking IPv4 fragment function

Command	Explanation
Global Mode	
[no] dosattack-check tcp-flags enable	Enable/disable checking TCP label function.
[no] dosattack-check ipv4-first-fragment enable	Enable/disable checking IPv4 fragment. This command has no effect when used separately, but if this function is not enabled, the switch will not drop the IPv4 fragment packet containing unauthorized TCP labels.

6.6.2.3 Anti Port Cheat Function Configuration Task Sequence

1. Enable the anti port cheat function

Command	Explanation
Global Mode	
[no] dosattack-check srcport-equal-dstport enable	Enable/disable the prevent-port-cheat function.

[no] dosattack-check ipv4-first-fragment enable	Enable/disable checking IPv4 fragment. This command has no effect when used separately, but if this function is not enabled, the switch will not drop the IPv4 fragment packet whose source port is equal to its destination port.

6.6.2.4 Prevent TCP Fragment Attack Function Configuration

Task Sequence

1. Enable the prevent TCP fragment attack function
2. Configure the minimum permitted TCP head length of the packet

Command	Explanation
Global Mode	
[no] dosattack-check tcp-fragment enable	Enable/disable the prevent TCP fragment attack function.
dosattack-check tcp-header *<size>*	Configure the minimum permitted TCP head length of the packet. This command has no effect when used separately, the user should enable the **dosattack-check tcp-fragment enable**.

6.6.2.5 Prevent ICMP Fragment Attack Function Configuration

Task Sequence

1. Enable the prevent ICMP fragment attack function
2. Configure the max permitted ICMPv4 net load length
3. Configure the max permitted ICMPv6 net load length

Command	Explanation
Global Mode	
[no] dosattack-check icmp-attacking enable	Enable/disable the prevent ICMP fragment attack function.
dosattack-check icmpv4-size *<size>*	Configure the max permitted ICMPv4 net load length. This command has not effect when used separately, the user have to enable the **dosattack-check icmp-attacking enable**.

dosattack-check icmpv6-size *<size>*	Configure the max permitted ICMPv6 net load length. This command has not effect when used separately, the user have to enable the **dosattack-check icmp-attacking enable**.

6.6.2.6 IPv4 Destination IP Check For Illegal Function

Configuration Task Sequence

1. Enable IPv4 destination IP checking for illegal function

Command	Explanation
Global Mode	
invalid-dip-drop {enable|disable}	Enable (disable) IPv4 destination IP checking for illegal function. Illegal destination IP includes X.X.X. 0, 127.X.X.X, 240.0.0.0~255.255.255.254.

6.6.3 Security Feature Example

Scenario:

The User has follows configuration requirements: the switch do not forward data packet whose source IP address is equal to the destination address, and those whose source port is equal to the destination port. Only the ping command with defaulted options is allowed within the IPv4 network, namely the ICMP request packet can not be fragmented and its net length is normally smaller than 100, enable IPv4 destination IP checking for illegal function.

Configuration procedure:

Switch(config)# dosattack-check srcip-equal-dstip enable
Switch(config)# dosattack-check srcport-equal-dstport enable
Switch(config)# dosattack-check ipv4-first-fragment enable
Switch(config)# dosattack-check icmp-attacking enable
Switch(config)# dosattack-check icmpV4-size 100
Switch(config)#invalid-dip-drop enable

6.7 TACACS+

6.7.1 Introduction to TACACS+

TACACS+ terminal access controller access control protocol is a protocol similar to the radius protocol for control the terminal access to the network. Three independent functions of Authentication, Authorization, Accounting are also available in this protocol. Compared with

RADIUS, the transmission layer of TACACS+ protocol is adopted with TCP protocol, further with the packet head (except for standard packet head) encryption, this protocol is of a more reliable transmission and encryption characteristics, and is more adapted to security control.

According to the characteristics of the TACACS+ (Version 1.78), we provide TACACS+ authentication function on the switch, when the user logs, such as telnet, the authentication of user name and password can be carried out with TACACS+.

6.7.2 TACACS+ Configuration Task List

1. Configure the TACACS+ authentication key
2. Configure the TACACS+ server
3. Configure the TACACS+ authentication timeout time
4. Configure the IP address of the RADIUS NAS

1. Configure the TACACS+ authentication key

Command	Explanation
Global Mode	
tacacs-server key {0 | 7}*<string>* **no tacacs-server key**	Configure the TACACS+ server key; the "**no tacacs-server key**" command deletes the key.

2. Configure TACACS+ server

Command	Explanation
Global Mode	
tacacs-server authentication host <*ip-address*> [port <*port-number*>] [timeout <*seconds*>] [key {0 | 7} <*string*>] [primary] **no tacacs-server authentication host <*ip-address*>**	Configure the IP address, listening port number, the value of timeout timer and the key string of the TACACS+ server; the no form of this command deletes the TACACS+ authentication server.

3. Configure the TACACS+ authentication timeout time

Command	Explanation
Global Mode	
tacacs-server timeout <*seconds*> **no tacacs-server timeout**	Configure the authentication timeout for the TACACS+ server, the "**no tacacs-server timeout**" command restores the default configuration.

4. Configure the IP address of the TACACS+ NAS

Command	Explanation
Global Mode	
tacacs-server nas-ipv4 <*ip-address*> **no tacacs-server nas-ipv4**	To configure the source IP address for the TACACS+ packets for the switch.

6.7.3 TACACS+ Scenarios Typical Examples

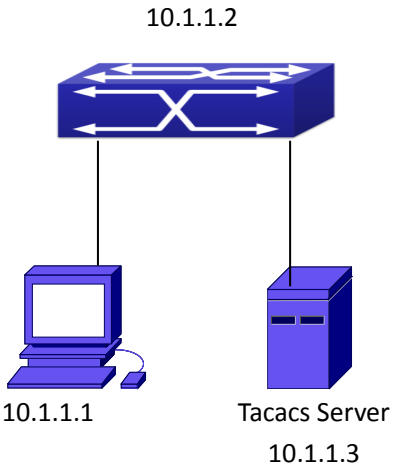


Figure 6-17 TACACS Configuration

A computer connects to a switch, of which the IP address is 10.1.1.2 and connected with a TACACS+ authentication server; IP address of the server is 10.1.1.3 and the authentication port is defaulted at 49, set telnet log on authentication of the switch as tacacs local, via using TACACS+ authentication server to achieve telnet user authentication.

Switch(config)#interface vlan 1
Switch(Config-if-vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-vlan1)#exit
Switch(config)#tacacs-server authentication host 10.1.1.3
Switch(config)#tacacs-server key test
Switch(config)#authentication line vty login tacacs

6.7.4 TACACS+ Troubleshooting

In configuring and using TACACS+, the TACACS+ may fail to authentication due to reasons such as physical connection failure or wrong configurations. The user should ensure the following:

- First good condition of the TACACS+ server physical connection.
- Second all interface and link protocols are in the UP state (use "**show interface**" command).
- Then ensure the TACACS+ key configured on the switch is in accordance with the one configured on TACACS+ server.
- Finally ensure to connect to the correct TACACS+ server.

6.8 RADIUS

6.8.1 Introduction to RADIUS

6.8.1.1 AAA and RADIUS Introduction

AAA is short for Authentication, Authorization and Accounting, it provide a consistency framework for the network management safely. According to the three functions of Authentication, Authorization, Accounting, the framework can meet the access control for the security network: which one can visit the network device, which access-level the user can have and the accounting for the network resource.

RADIUS (Remote Authentication Dial in User Service), is a kind of distributed and client/server protocol for information exchange. The RADIUS client is usually used on network appliance to implement AAA in cooperation with 802.1x protocol. The RADIUS server maintains the database for AAA, and communicates with the RADIUS client through RADIUS protocol. The RADIUS protocol is the most common used protocol in the AAA framework.

6.8.1.2 Message structure for RADIUS

The RADIUS protocol uses UDP to deliver protocol packets. The packet format is shown as below.

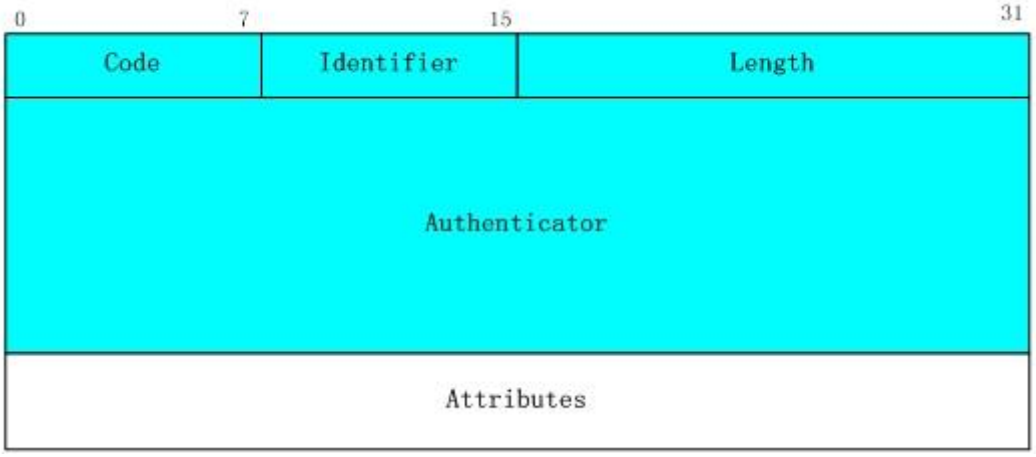


Figure 6-18 Message structure for RADIUS

Code field(1octets): is the type of the RADIUS packet. Available value for the Code field is show as below:

1 Access-Request
2 Access-Accept
3 Access-Reject
4 Accounting-Request
5 Accounting-Response

11 Access-Challenge

Identifier field (1 octet): Identifier for the request and answer packets.

Length field (2 octets): The length of the overall RADIUS packet, including Code, Identifier, Length, Authenticator and Attributes

Authenticator field (16 octets): used for validation of the packets received from the RADIUS server. Or it can be used to carry encrypted passwords. This field falls into two kinds: the Request Authenticator and the Response Authenticator.

Attribute field: used to carry detailed information about AAA. An Attribute value is formed by Type, Length, and Value fields.

☞ Type field (1 octet), the type of the attribute value, which is shown as below:

Property	Type of property	Property	Type of property
1	User-Name	23	Framed-IPX-Network
2	User-Password	24	State
3	CHAP-Password	25	Class
4	NAS-IP-Address	26	Vendor-Specific
5	NAS-Port	27	Session-Timeout
6	Service-Type	28	Idle-Timeout
7	Framed-Protocol	29	Termination-Action
8	Framed-IP-Address	30	Called-Station-Id
9	Framed-IP-Netmask	31	Calling-Station-Id
10	Framed-Routing	32	NAS-Identifier
11	Filter-Id	33	Proxy-State
12	Framed-MTU	34	Login-LAT-Service
13	Framed-Compression	35	Login-LAT-Node
14	Login-IP-Host	36	Login-LAT-Group
15	Login-Service	37	Framed-AppleTalk-Link
16	Login-TCP-Port	38	Framed-AppleTalk-Network
17	(unassigned)	39	Framed-AppleTalk-Zone
18	Reply-Message	40-59	(reserved for accounting)
19	Callback-Number	60	CHAP-Challenge
20	Callback-Id	61	NAS-Port-Type
21	(unassigned)	62	Port-Limit
22	Framed-Route	63	Login-LAT-Port

☞ Length field (1 octet), the length in octets of the attribute including Type, Length and Value fields.

☞ Value field, value of the attribute whose content and format is determined by the type and length of the attribute.

6.8.2 RADIUS Configuration Task List

1. Enable the authentication and accounting function
2. Configure the RADIUS authentication key
3. Configure the RADIUS server
4. Configure the parameter of the RADIUS service
5. Configure the IP address of the RADIUS NAS

1. Enable the authentication and accounting function

Command	Explanation
Global Mode	
aaa enable **no aaa enable**	To enable the AAA authentication function. The no form of this command will disable the AAA authentication function.
aaa-accounting enable **no aaa-accounting enable**	To enable AAA accounting. The no form of this command will disable AAA accounting.
aaa-accounting update {enable/disable}	Enable or disable the update accounting function.

2. Configure the RADIUS authentication key

Command	Explanation
Global Mode	
radius-server key {0 | 7} <*string*> **no radius-server key**	To configure the encryption key for the RADIUS server. The no form of this command will remove the configured key.

3. Configure the RADIUS server

Command	Explanation
Global Mode	
radius-server authentication host {<*ipv4-address*> | <*ipv6-address*>} [port <*port-number*>] [key *{0 | 7}* <string>] [primary] [access-mode {dot1x | telnet}] **no radius-server authentication host {<*ipv4-address*> | <*ipv6-address*>**	Specifies the IPv4/IPv6 address and the port number, whether be primary server for RADIUS accounting server; the no command deletes the RADIUS accounting server.
radius-server accounting host {<ipv4-address> | <ipv6-address>} [port <port-number>] [key {*0* | 7} <string>] [primary] **no radius-server accounting host {<ipv4-address> / <ipv6-address>}**	Specifies the IPv4/IPv6 address and the port number, whether be primary server for RADIUS accounting server; the no command deletes the RADIUS accounting server.

4. Configure the parameter of the RADIUS service

Command	Explanation
Global Mode	
radius-server dead-time *<minutes>* **no radius-server dead-time**	To configure the interval that the RADIUS becomes available after it is down. The no form of this command will restore the default configuration.
radius-server retransmit *<retries>* **no radius-server retransmit**	To configure retry times for the RADIUS packets. The no form of this command restores the default configuration.
radius-server timeout *<seconds>* **no radius-server timeout**	To configure the timeout value for the RADIUS server. The no form of this command will restore the default configuration.
radius-server accounting-interim-update timeout *<seconds>* **no radius-server accounting-interim-update timeout**	To configure the update interval for accounting. The no form of this command will restore the default configuration.

5. Configure the IP address of the RADIUS NAS

Command	Explanation
Global Mode	
radius nas-ipv4 *<ip-address>* **no radius nas-ipv4**	To configure the source IP address for the RADIUS packets for the switch.
radius nas-ipv6 *<ipv6-address>* **no radius nas-ipv6**	To configure the source IPv6 address for the RADIUS packets for the switch.

6.8.3 RADIUS Typical Examples

6.8.3.1 IPv4 Radius Example

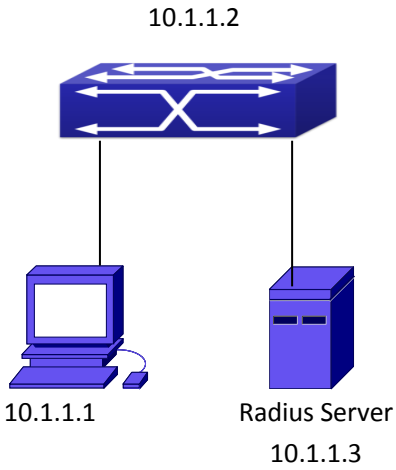


Figure 6-19 The Topology of IEEE802.1x configuration

A computer connects to a switch, of which the IP address is 10.1.1.2 and connected with a RADIUS authentication server without Ethernet1/0/2; IP address of the server is 10.1.1.3 and the authentication port is defaulted at 1812, accounting port is defaulted at 1813.

Configure steps as below:

Switch(config)#interface vlan 1
Switch(Config-if-vlan1)#ip address 10.1.1.2 255.255.255.0
Switch(Config-if-vlan1)#exit
Switch(config)#radius-server authentication host 10.1.1.3
Switch(config)#radius-server accounting host 10.1.1.3
Switch(config)#radius-server key test
Switch(config)#aaa enable
Switch(config)#aaa-accounting enable

6.8.3.2 IPv6 RadiusExample

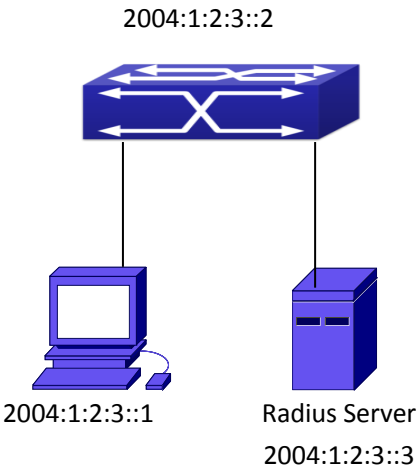


Figure 6-20 The Topology of IPv6 Radius configuration

A computer connects to a switch, of which the IP address is 2004:1:2:3::2 and connected with a RADIUS authentication server without Ethernet1/0/2; IP address of the server is 2004:1:2:3::3 and the authentication port is defaulted at 1812, accounting port is defaulted at 1813.

Configure steps as below:
Switch(config)#interface vlan 1
Switch(Config-if-vlan1)#ipv6 address 2004:1:2:3::2/64
Switch(Config-if-vlan1)#exit
Switch(config)#radius-server authentication host 2004:1:2:3::3
Switch(config)#radius-server accounting host 2004:1:2:3::3
Switch(config)#radius-server key test
Switch(config)#aaa enable
Switch(config)#aaa-accounting enable

6.8.4 RADIUS Troubleshooting

In configuring and using RADIUS, the RADIUS may fail to authentication due to reasons such as physical connection failure or wrong configurations. The user should ensure the following:

- ☞ First make sure good condition of the RADIUS server physical connection
- ☞ Second all interface and link protocols are in the UP state (use "**show interface**" command)
- ☞ Then ensure the RADIUS key configured on the switch is in accordance with the one configured on RADIUS server
- ☞ Finally ensure to connect to the correct RADIUS server

If the RADIUS authentication problem remains unsolved, please use **debug aaa** and other debugging command and copy the DEBUG message within 3 minutes, send the recorded message to the technical server center of our company.

6.9 SSL

6.9.1 Introduction to SSL

As the computer networking technology spreads, the security of the network has been taking more and more important impact on the availability and the usability of the networking application. The network security has become one of the greatest barriers of modern networking applications.

To protect sensitive data transferred through Web, Netscape introduced the Secure Socket

Layer – SSL protocol, for its Web browser. Up till now, SSL 2.0 and 3.0 has been released. SSL 2.0 is obsolete because of security problems, and it is not supported on the switches of Network. The SSL protocol uses the public-key encryption, and has become the industry standard for secure communication on internet for Web browsing. The Web browser integrates HTTP and SSL to realize secure communication.

SSL is a safety protocol to protect private data transmission on the Internet. SSL protocols are designed for secure transmission between the client and the server, and authentication both at the server sides and optional client. SSL protocols must build on reliable transport layer (such as TCP). SSL protocols are independent for application layer. Some protocols such as HTTP, FTP, TELNET and so on, can build on SSL protocols transparently. The SSL protocol negotiates for the encryption algorithm, the encryption key and the server authentication before data is transmitted. Ever since the negotiation is done, all the data being transferred will be encrypted.

Via above introduction, the security channel is provided by SSL protocols have below three characteristics:

- ☞ Privacy. First they encrypt the suite through negotiation, then all the messages be encrypted.
- ☞ Affirmation. Though the client authentication of the conversational is optional, but the server is always authenticated.
- ☞ Reliability. The message integrality inspect is included in the sending message (use MAC).

6.9.1.1 Basic Element of SSL

The basic strategy of SSL provides a safety channel for random application data forwarding between two communication programs. In theory, SSL connect is similar with encrypt TCP connect. The position of SSL protocol is under application layer and on the TCP. If the mechanism of the data forwarding in the lower layer is reliable, the data read-in the network will be forwarded to the other program in sequence, lose packet and re-forwarding will not appear. A lot of transmission protocols can provide such kind of service in theory, but in actual application, SSL is almost running on TCP, and not running on UDP and IP directly.

When web function is running on the switch and client visit our web site through the internet browser, we can use SSL function. The communication between client and switch through SSL connect can improve the security.

Firstly, SSL should be enabled on the switch. When the client tries to access the switch through https method, a SSL session will be set up between the switch and the client. When the SSL session has been set up, all the data transmission in the application layer will be encrypted.

SSL handshake is done when the SSL session is being set up. The switch should be able to provide certification keys. Currently the keys provided by the switch are not the formal certification keys issued by official authentic, but the private certification keys generated by SSL software under Linux which may not be recognized by the web browser. With regard to the switch application, it is not necessary to apply for a formal SSL certification key. A private certification key is enough to make the communication safe between the users and the switch. Currently it is not required that the client is able to check the validation of the certification key. The encryption key and the encryption method should be negotiated during the handshake period of the session which will be then used for data encryption.

SSL session handshake process:

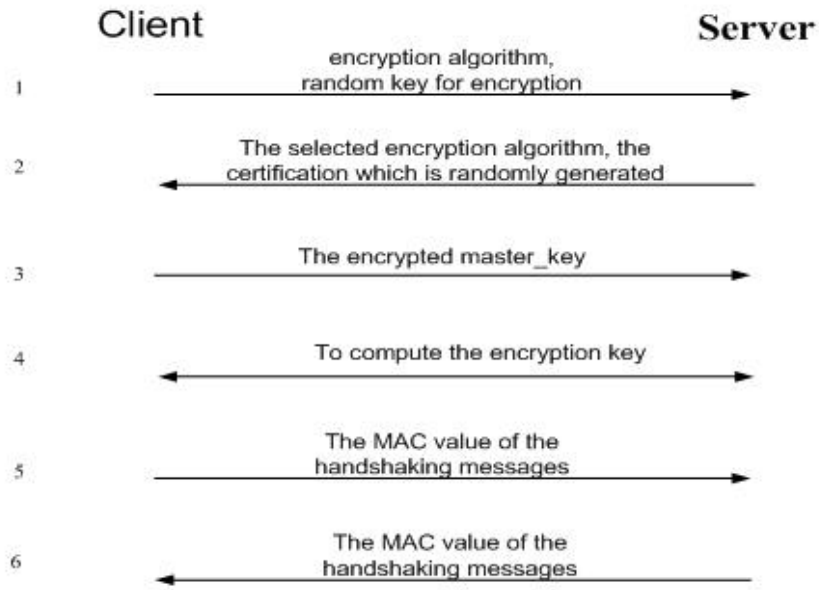


Figure 6-201 SSL session handshake process

6.9.2 SSL Configuration Task List

1. Enable/disable SSL function
2. Configure/delete port number by SSL used
3. Configure/delete secure cipher suite by SSL used
4. Maintenance and diagnose for the SSL function

1. Enable/disable SSL function

Command	Explanation
Global Mode	
ip http secure-server **no ip http secure-server**	Enable/disable SSL function.

2. Configure/delete port number by SSL used

Command	Explanation
Global Mode	
ip http secure-port <*port-number*> **no ip http secure-port**	Configure port number by SSL used, the"**no ip http secure-port**" command deletes the port number.

3. Configure/delete secure cipher suite by SSL used

Command	Explanation
Global Mode	
ip http secure-ciphersuite {des-cbc3-sha|rc4-128-sha| des-cbc-sha} **no ip http secure-ciphersuite**	Configure/delete secure cipher suite by SSL used.

4. Maintenance and diagnose for the SSL function

Command	Explanation
Admin Mode or Configuration Mode	
show ip http secure-server status	Show the configured SSL information.
debug ssl **no debug ssl**	Open/close the DEBUG for SSL function.

6.9.3 SSL Typical Example

When the Web function is enabled on the switch, SSL can be configured for users to access the web interface on the switch. If the SSL has been configured, communication between the client and the switch will be encrypted through SSL for safety.

Firstly, SSL should be enabled on the switch. When the client tries to access the switch through https method, a SSL session will be set up between the switch and the client. When the SSL session has been set up, all the data transmission in the application layer will be encrypted.

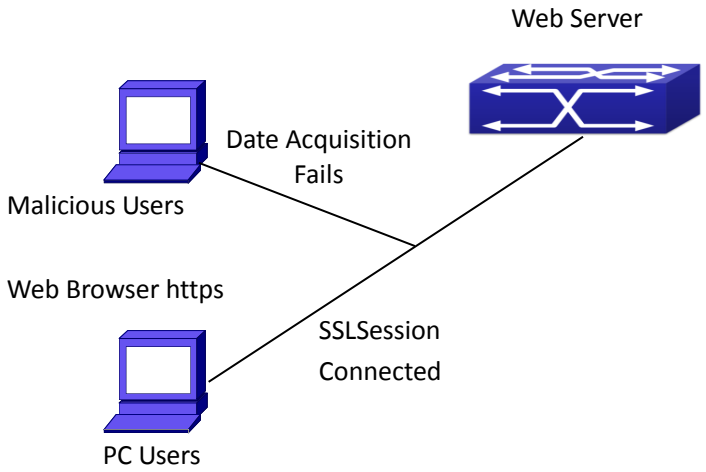


Figure 6-212

Configuration on the switch:
Switch(config)# ip http secure-server
Switch(config)# ip http secure-port 1025
Switch(config)# ip http secure-ciphersuite rc4-128-sha

6.9.4 SSL Troubleshooting

In configuring and using SSL, the SSL function may fail due to reasons such as physical connection failure or wrong configurations. The user should ensure the following:

- First good condition of the physical connection;
- Second all interface and link protocols are in the UP state (use “show interface” command);
- Then, make sure SSL function is enabled (use ip http secure-server command);
- Don’t use the default port number if configured port number, pay attention to the port number when input the web wide;

- ☞ If SSL is enabled, SSL should be restarted after changes on the port configuration and encryption configuration;
- ☞ IE 7.0 or above should be used for use of des-cbc-sha;
- ☞ If the SSL problems remain unsolved after above try, please use debug SSL and other debugging command and copy the DEBUG message within 3 minutes, send the recorded message to technical server center of our company.

6.10 VLAN-ACL

6.10.1 Introduction to VLAN-ACL

The user can configure ACL policy to VLAN to implement the accessing control of all ports in VLAN, and VLAN-ACL enables the user to expediently manage the network. The user only needs to configure ACL policy in VLAN, the corresponding ACL action can takes effect on all member ports of VLAN, but it does not need to solely configure on each member port.

When VLAN ACL and Port ACL are configured at the same time, the principle of denying firstly is used. When the packets match VLAN ACL and Port ACL at the same time, as long as one rule is drop, then the final action is drop.

Egress ACL can implement the filtering of the packets on egress and ingress direction, the packets match the specific rules can be allowed or denied. ACL can support IP ACL, MAC ACL, MAC-IP ACL, IPv6 ACL. Ingress direction of VLAN can bind four kinds of ACL at the same time, there are four resources on egress direction of VLAN, IP ACL and MAC ACL engage one resource severally, MAC-IP ACL and IPv6 ACL engage two resources severally, so egress direction of VLAN can not bind four kinds of ACL at the same time. When binding three kinds of ACL at the same time, it should be the types of IP, MAC, MAC-IP or IP, MAC, IPv6. When binding two kinds of ACL at the same time, any combination of ACL type is valid. Each type can only apply one on a VLAN.

6.10.2 VLAN-ACL Configuration Task List

1. Configure VLAN-ACL of IP type
2. Configure VLAN-ACL of MAC type
3. Configure VLAN-ACL of MAC-IP
4. Configure VLAN-ACL of IPv6 type
5. Show configuration and statistic information of VLAN-ACL
6. Clear statistic information of VLAN-ACL

1. Configure VLAN-ACL of IP type

Command	Explanation
Global mode	

vacl ip access-group {<1-299> | WORD} {in | out} [traffic-statistic] vlan WORD **no vacl ip access-group {<1-299> | WORD} {in | out} vlan WORD**	Configure or delete IP VLAN-ACL.

2. Configure VLAN-ACL of MAC type

Command	Explanation
Global mode	
vacl mac access-group {<700-1199> | WORD} {in | out} [traffic-statistic] vlan WORD **no vacl mac access-group {<700-1199> | WORD} {in | out} vlan WORD**	Configure or delete MAC VLAN-ACL.

3. Configure VLAN-ACL of MAC-IP

Command	Explanation
Global mode	
vacl mac-ip access-group {<3100-3299> | WORD} {in | out} [traffic-statistic] vlan WORD **no vacl mac-ip access-group {<3100-3299> | WORD} {in | out} vlan WORD**	Configure or delete MAC-IP VLAN-ACL.

4. Configure VLAN-ACL of IPv6 type

Command	Explanation
Global mode	
vacl ipv6 access-group (<500-699> | WORD) {in | out} (traffic-statistic|) vlan WORD **no ipv6 access-group {<500-699> | WORD} {in | out} vlan WORD**	Configure or delete IPv6 VLAN-ACL.

5. Show configuration and statistic information of VLAN-ACL

Command	Explanation
Admin mode	
show vacl [in | out] vlan [<*vlan-id*>]	Show the configuration and the statistic information of VACL.

6. Clear statistic information of VLAN-ACL

Command	Explanation
Admin mode	
clear vacl [in | out] statistic vlan [<*vlan-id*>]	Clear the statistic information of VACL.

6.10.3 VLAN-ACL Configuration Example

A company's network configuration is as follows, all departments are divided by different VLANs, technique department is Vlan1, finance department is Vlan2. It is required that technique department can access the outside network at timeout, but finance department are not allowed to access the outside network at any time for the security. Then the following policies are configured:

- ☞ Set the policy VACL_A for technique department. At timeout they can access the outside network, the rule as permit, but other times the rule as deny, and the policy is applied to Vlan1.
- ☞ Set the policy VACL_B of ACL for finance department. At any time they can not access the outside network, but can access the inside network with no limitation, and apply the policy to Vlan2.

Network environment is shown as below:

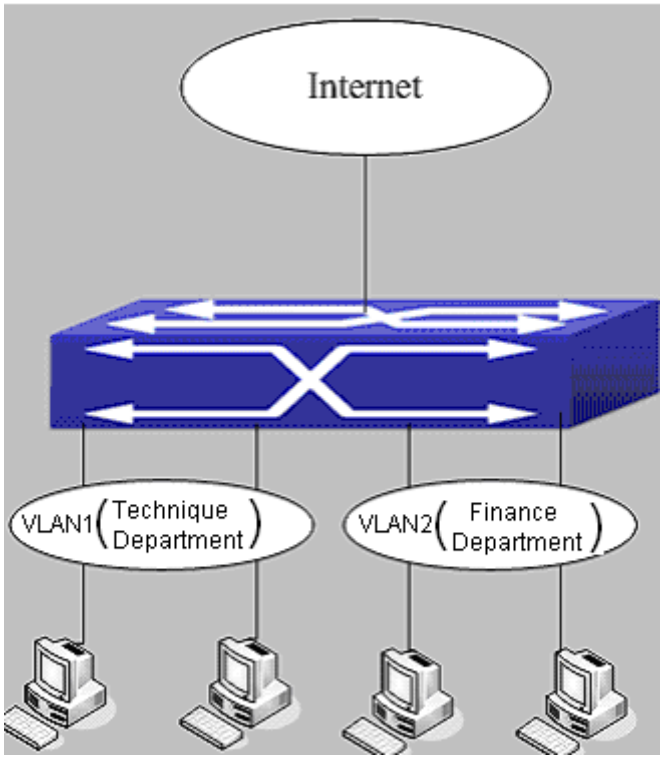


Figure 6-23 VLAN-ACL configuration example

Configuration example:

1) First, configure a timerange, the valid time is the working hours of working day:

Switch(config)#time-range t1

Switch(config-time-range-t1)#periodic weekdays 9:00:00 to 12:00:00

Switch(config-time-range-t1)#periodic weekdays 13:00:00 to 18:00:00

2) Configure the extended acl_a of IP, at working hours it only allows to access the resource within the internal network (such as 192.168.0.255).

Switch(config)# ip access-list extended vacl_a

Switch(config-ip-ext-nacl-vacl_a)# permit ip any-source 192.168.0.0 0.0.0.255 time-range t1

Switch(config-ip-ext-nacl-vacl_a)# deny ip any-source any-destination time-range t1

3) Configure the extended acl_b of IP, at any time it only allows to access resource within the

internal network (such as 192.168.1.255).

```
Switch(config)#ip access-list extended vacl_b
Switch(config-ip-ext-nacl-vacl_a)# permit ip any-source 192.168.1.0 0.0.0.255
Switch(config-ip-ext-nacl-vacl_a)# deny ip any-source any-destination
```

4) Apply the configuration to VLAN

```
Switch(config)#vacl ip access-group vacl_a in vlan 1
Switch(config)#vacl ip access-group vacl_b in vlan 2
```

6.10.4 VLAN-ACL Troubleshooting

When VLAN ACL and Port ACL are configured at the same time, the principle of denying firstly is used. When the packets match VLAN ACL and Port ACL at the same time, as long as one rule is drop, then the final action is drop.

☞ Each ACL of different types can only apply one on a VLAN, such as the basic IP ACL, each VLAN can applies one only.

6.11 Captive Portal Authentication

6.11.1 Captive Portal Authentication Configuration

6.11.1.1 Introduction to Captive Portal Authentication

The authentication function is a way to manage and control the network resources for users. Authentication function memories the client authentication information in the authentication server according to a certain principles. When a user needs to use the network resources, the function of captive portal will redirect the network request of user to the authentication server, and then the user needs to provide allowed username, password and other information, the authentication server will judge the information of user and decide whether the user can be allowed to use the network resources. The switch in authentication function plays a role of communicating the user and the authentication server. Through the switch configuration, it enables the user could connect and communicate with the authentication server, and the server will analyze the data and provide the corresponding feedback to the user. Authentication function based on the redirection function.

Redirection is a function of re-connecting the original request to a predetermined site and continous to operate. The function is when the AP receives a client request, then transfer the client request to a predetermined address, after the operation of the client and the redirected address, in order to complete certain functions and operations. This operation can achieve the aim to manage and monitor the user. Client redirected to portal authentication interface, requiring the user to fill in the username and password, only when the username and password

pass the certification and they can use the network resources. Portal authentication can achieve different control strategies for different types of users.

The portal server function is a way to configure different external portal server for different CP configuration. When network bind different CP configuration, has configured different portal server, it will launch the redirect page through their binding portal server. You can configure up to 10 external portal servers. Each CP configuration can bind one portal server.

6.11.1.2 Captive Portal Authentication Configuration

Authentication function task list is as below:

1. Enable/disable captive portal authentication function
2. Configure captive portal redirect function
   1) Configure or delete the portal server name
   2) Configure the user instance
   3) Configure redirect address
   4) Configure the redirect url-head
   5) Configure/delete radius server name
   6) Configure to bind or relieve the portal server
   7) Configure the url to carry the parameter of ac-name
   8) Configure the url to carry the parameter of ssid
   9) Configure the url to carry the parameter of nas-ip
3. Configure AAA function
   1) Enable/stop AAA function
   2) Configure RADIUS authentication server group name
4. Configure RADIUS authentication server
   1) Configure RADIUS server key
   2) Configure RADIUS authentication server address
5. Bind the portal rule to the port

1. Enable/disable captive portal authentication function

Command	Explanation
Captive Portal Configuration Mode	
enable **disable**	Enable/disable captive portal function globally.

2. Configure captive portal redirect function

Command	Explanation
Captive Portal Configuration Mode	

external portal-server server-name <*name*> {ipv4 | ipv6} <*ipaddr*> [port <*1-65535*>] **no external portal-server {ipv4 | ipv6}server-name <*name*>**	Configure/delete external portal server.
nas-ip4 <A.B.C.D>	Configure the nas-ip address.
configuration <*cp-id*> **no configuration <*cp-id*>**	Configure/delete portal routine of different types of users. 10 kinds of routines can be configured.
Captive Portal Instance Configuration Mode	
redirect url-head <word> **no redirect url-head**	Configure the redirect url-head including transmission protocol, host name, port and path. The no command deletes the configuration.
radius-auth-server <*server-name*> **no radius-auth-server**	Configure/delete authentication server name.
portal-server {ipv4 | ipv6} <*name*> **no portal-server {ipv4 | ipv6}**	Bind/unbind portal server name.
enable **disable**	Enable/disable a portal routine.
ac-name <word> **no ac-name**	Configure the url to carry the parameter of acname. The no command deletes it.
redirect attribute ssid name <word> **no redirect attribute ssid name**	Configure the url to carry the parameter of ssid. The no command recovers the ssid to be the default value.
redirect attribute nas-ip enable **no redirect attribute nas-ip enable**	Configure the url to carry the parameter of nas-ip. The no command disables this function.
redirect attribute nas-ip name <word> **no redirect attribute nas-ip name**	Configure the name of the parameter of nas-ip carried in url. The no command recovers the name to be the default value.

3. Configure AAA function

Command	Explanation
Global Mode	
aaa enable **no aaa enable**	Enable/stop the AAA function of a captive portal routine.

aaa group server radius <*word*> **no aaa group server radius <*word*>**	Configure/delete RADIUS name of AAA function.

4. Configure RADIUS authentication server

Command	Explanation
Global Mode	
radius-server key <*word*> **no radius-server key**	Configure/delete RADIUS server key.
radius-server authentication host <*A.B.C.D*> **no radius-server authentication host <*A.B.C.D*>**	Configure/delete RADIUS authentication server address.

5. Bind the portal rule to the port

Command	Explanation
Config Mode	
vlan-pool <1-255> <WORD> **no vlan-pool <1-255>**	Configure or delete the vlan pool. (optional)
Port Mode	
portal enable configuration <id> [vlan-pool WORD] **no portal enable [vlan-pool WORD]**	Enable the port portal authentication function and appoint the instance number which is bound to the port. It can also appoint which vlan to be enabled portal. If the vlan is not appointed, all vlan is enabled portal as default. The no command disables the portal authentication function of the port. (necessary)

6.11.1.3 Captive Portal Authentication Examples

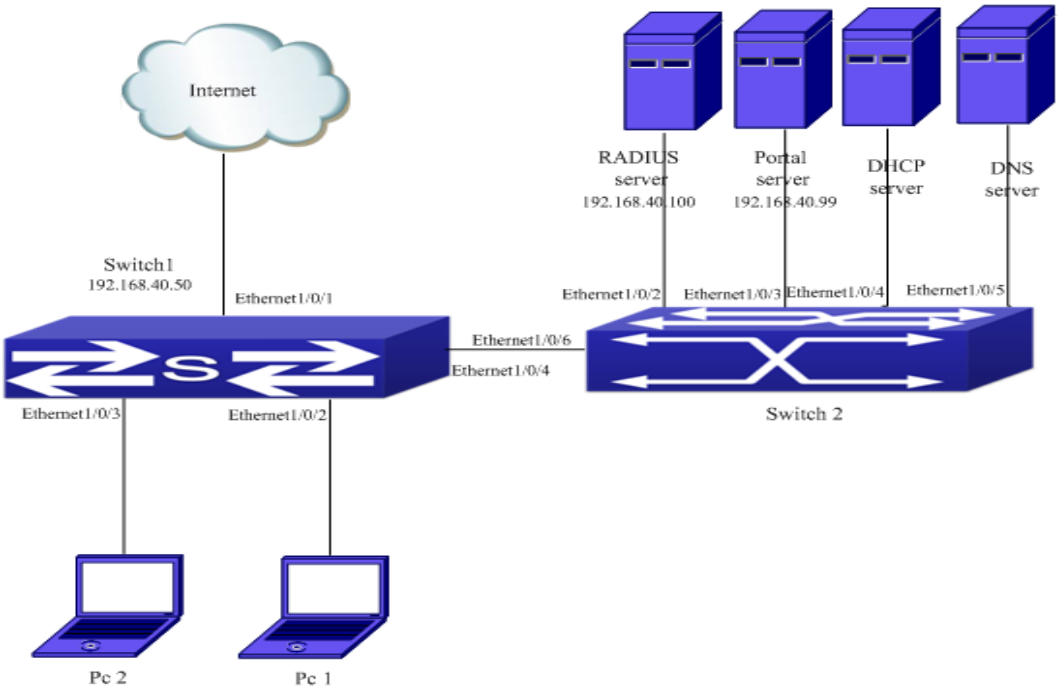


Figure 6-24 authentication function configuration

As shown above, pc1 is the terminal client, there is the http browser but not the 802.1x authentication client, pc1 wants to access the network through the portal authentication.

The switch1 is the accessing device with the configured accounting server addressas the IP and port of the radius server, and it is enabled the accounting function. The Ethernet1/0/2 is connected to pc1, the port is enabled portal authentication function, and the redirection address is configured as the IP and port of the portal server. So the Ethernet1/0/2 forbids all the traffic and only allows the dhcp/dns/arp packets.

The switch2 is the convergence switch, the Ethernet1/0/2 is communicated to the radius server, the Ethernet1/0/3 is communicated to the portal server. The address of the radius server is 192.168.40.100, and the address of the portal server is 192.168.40.99. The Ethernet1/0/4 is connected to DHCP server and the Ethernet1/0/5 is connected to DNS server. The Ethernet1/0/6 is the trunk port and connected to the Ethernet1/0/4 of the trunk port of switch1.

Configure the radius server:
switch (config)#radius-server key 0 test
switch (config)#aaa group server radius radius_aaa_1
switch (config-sg-radius)# server 192.168.40.100

The configuration of global authentication:
Switch(config)#interface vlan 1
Switch(config-if-vlan1)#ip address 192.168.40.50 255.255.255.0
Switch(config)#free-resource 1 destination ipv4 192.168.40.99/32
Switch(config)# vlan-pool 1 2-10

Configure the portal function and portal server under the portal instance:

```
Switch (config)#captive-portal
Switch (config-cp)#enable
Switch(config-cp)# nas-ipv4 192.168.40.50
Switch(config-cp)# external portal-server server-name abc ipv4 192.168.40.99
Switch (config-cp)# configuration 1
Switch (config-cp-instance)#name helix4
Switch (config-cp-instance)#radius-auth-server abc99
Switch (config-cp-instance)# redirect attribute nas-ip enable
Switch (config-cp-instance)#redirect attribute nas-ip name kk
Switch (config-cp-instance)#ac-name helix4
Switch (config-cp-instance)#redirect url-head http://192.168.40.99/a70.htm
Switch (config-cp-instance) # portal-server ipv4 abc
```

Enable the portal function on the port:

```
Switch (config-cp-instance) #interface ethernet 1/0/2 vlan add 1
```

6.11.1.4 Captive Portal Authentication Troubleshooting

Encounter problems when using the redirection function, please check whether the reasons are as follows:

- ☞ Whether launched the captive portal function and opened the portal configuration switch. Both the captive portal and the portal configuration should be open; otherwise, the captive portal function will not work, the client also cannot be redirect to the specified page.
- ☞ The authentication server name of AAA module is same to the configured authentication name of captive portal.
- ☞ Whether the port which connects the pc and switch is enabled portal authentication function.

6.11.2 Accounting Function Configuration

6.11.2.1 Introduction to Accounting Function

The accounting function is used to monitoring and accounting users who using the network resources. Client is unable to access the network resources before pass the captive portal authentication, only through the portal authentication to access network resources. Define user's session duration to control the use of network resources time and flow of information.

6.11.2.2 Accounting Function Configuration

Accounting function configuration task list is as below:

1. Configure RADIUS accounting server
   1) Configure/delete accounting server address

2. Configure AAA accounting function
   1) Enable/disable accounting service function
3. Configure captive portal accounting function
   1) Block/unblock portal function
   2) Configure/delete captive portal configuration name
   3) Enable/disable captive portal accounting function
   4) Configure/delete captive portal accounting server name
   5) Configure/delete captive portal session time

1. Configure RADIUS accounting server

Command	Explanation
Global Mode	
radius-server accounting host <*A.B.C.D*> **no radius-server accounting host <*A.B.C.D*>**	Configure/delete accounting server address

2. Configure AAA accounting function

Command	Explanation
Global Mode	
aaa-accounting enable **no aaa-accounting**	Enable/disable accounting service function

3. Configure captive portal accounting function

Command	Explanation
Captive Portal Configuration Mode	
block **no block**	Block/unblock portal function
name <*word*> **no name**	Configure/delete captive portal configuration name
radius accounting **no aaa-accounting**	Enable/disable captive portal accounting function
radius-acct-server <*word*> **no radius-acct-server**	Configure/delete captive portal accounting server name
session-timeout <*0-86400*> **session-timeout**	Configure/delete captive portal session time

6.11.2.3 Accounting Function Examples

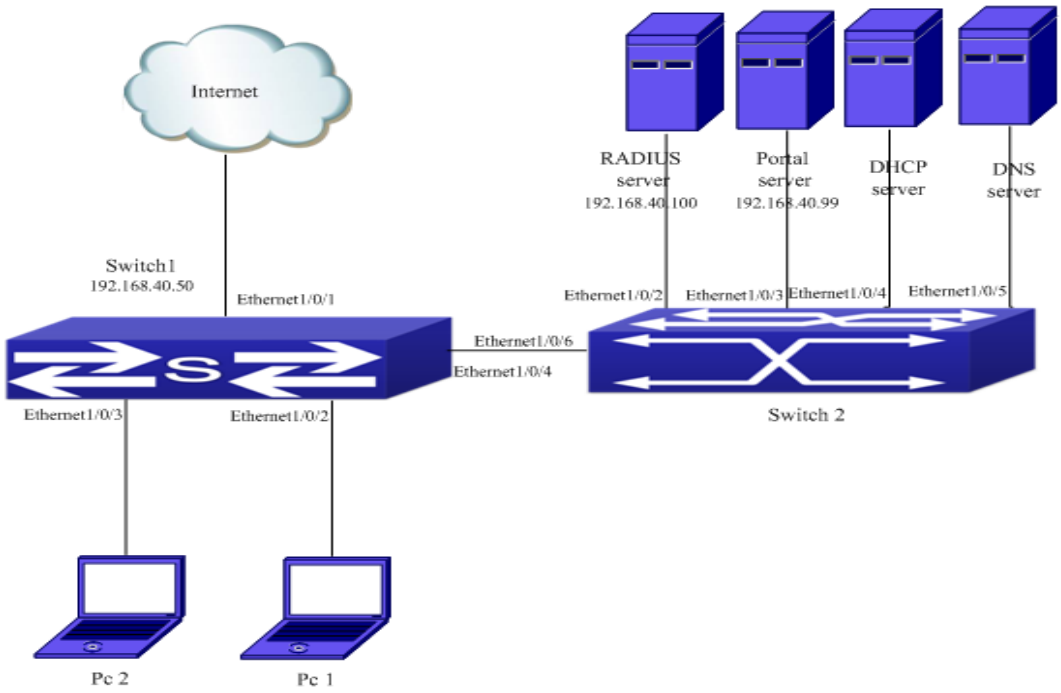


Figure 6-25 accounting function configuration

1. Configure the AAA accounting function on switch1.

AAA configuration of Switch1:

switch 1(config)# aaa enable

switch 1(config)# aaa-accounting enable

switch 1(config)# radius-server accounting host 192.168.40.100

switch1 (config)#radius-server key 0 test

switch1 (config)#aaa group server radius abc99

switch (config-sg-radius)# server 192.168.40.100

2. Configure the captive portal accounting function on switch1.

Switch 1(config)#captive-portal

Switch 1(config-cp)#enable

Switch1 (config-cp)# configuration 1

Switch1 (config-cp-instance)#radius accounting

Switch1 (config-cp-instance)# radius-acct-server abc99

6.11.2.4 Accounting Function Troubleshooting

Encounter problems when using the accounting function, please check whether the reasons are as follows:

☞ Whether launched the captive portal function and opened the portal configuration switch. Both the captive portal and the portal configuration should be open; otherwise, the captive portal function will not work, the client also cannot be redirect to the specified page.

☞ The authentication server name of AAA module is same to the configured authentication name of captive portal.

6.11.3 Free-resource Configuration

6.11.3.1 Introduction to Free-resource

Free-resource function is a method of captive portal module to achieve access the free resources rule. By configuring the free-resource rules, one can makes certain the client directly access the specific network resources without going through the portal authentication.

6.11.3.2 Free-resource Configuration

1. Configure the free-resource rule

Command	Explanation
Config Mode	
free-resource destination {ipv4 A.B.C.D/M |ipv6 X:X:X:X/M} **[no] free-resource destination {ipv4 A.B.C.D/M |ipv6 X:X:X:X/M}**	Configure or delete the free-resource rule.

6.11.3.3 Free-resource Examples

Case：

Set up an environment as shown below. The IP is the address segment for the Client1, and the Destination IP is the address segment for client who wants to access the resources. Appoint RADIUS server 1 as the authentication server, client1 and client2 can access the free-resource of 3.1.1.0/24 and will not be redirected to the authentication server.

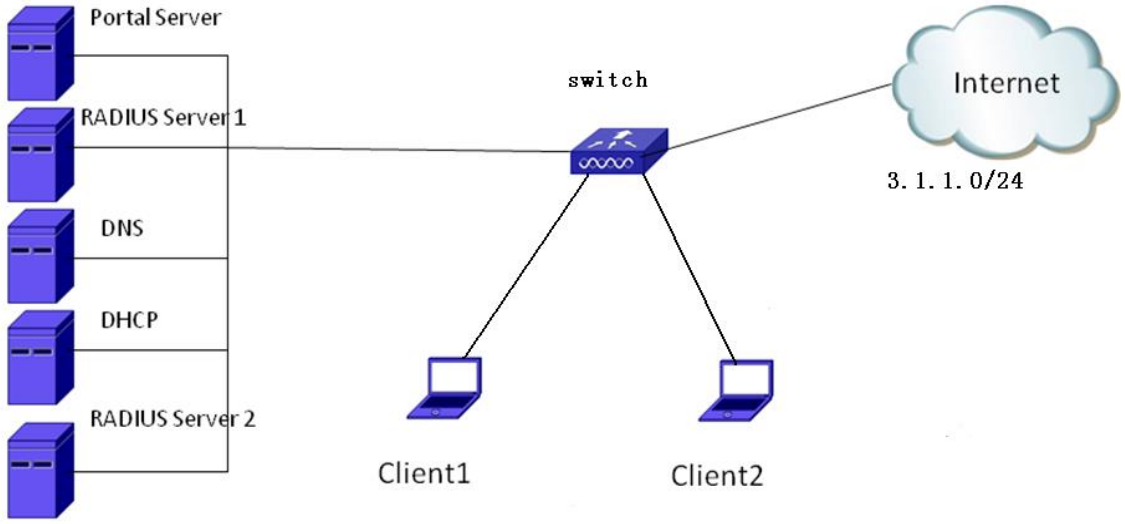


Figure 6-26 multi-portal servers function configuration

Configuration steps:

Switch1(config-)# free-resource destination ipv4 3.1.1.0/24

6.11.3.4 Free-resource Troubleshooting

Encounter problems when using the redirection function, please check whether the reasons are as follows:

- ☞ Whether launched the captive portal function and opened the portal configuration switch. Both the captive portal and the portal configuration should be open; otherwise, the captive portal function will not work, the client also cannot be redirect to the specified page.
- ☞ Whether the port which connects the client and switch is enabled portal rule.

6.11.4 Authentication White-list Configuration

6.11.4.1 Introduction to Authentication White-list

Authentication white-list is used for some special users in the network. The administrator can set permission to allow the users to connect to the network to use network resources without authentication, but the administrator needs to get the user's mac address. At the same time the user who has the permission to use network resources do not need to billing. So the user belongs to the advanced user.

6.11.4.2 Authentication White-list Configuration

1. Configure user mac with Authentication white-list function purview

Command	Explanation
Config Mode	
free mac < MACADD><MACMASK> **no free mac < MACADD><MACMASK>**	Configure or delete the mac address without needing to authenticate.

6.11.4.3 Authentication White-list Examples

Case:

As shown below, client1 and client2 are the terminal clients; the port connected to the switch is enabled portal authentication. But these two clients are the advanced users, they need no authentication to access the network.

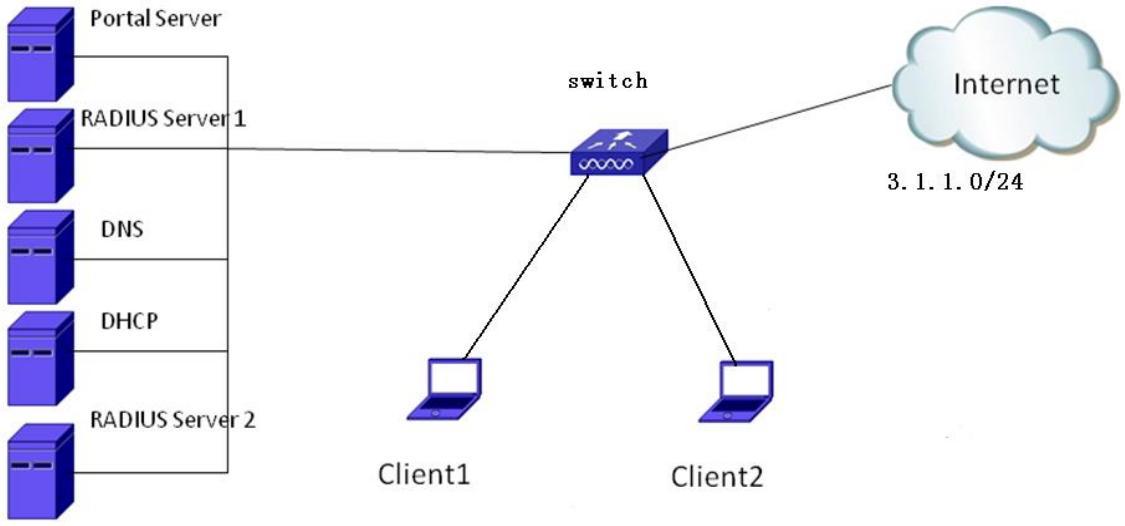


Figure 6-27 Authentication White-list function configuration

Configuration steps:

Configure the authentication white-list for pc1 and pc2.

Switch1 (config)#free-mac 68-74-7f-29-76-04 ff-ff-ff-ff-ff-ff

Switch1 (config)#free-mac 00-03-0f-11-11-11 ff-ff-ff-ff-ff-ff

6.11.4.4 Authentication White-list Troubleshooting

When encountered problems in the process of using the Authentication White-list function, please check whether the reasons are as follows:

- Whether the free mac and the client mac is matching.
- Whether the port which connects the client and switch is enabled the portal rule.

6.11.5 Automatic Page Pushing after Successful Authentication (it is not supported currently)

6.11.5.1 Introduction to Automatic Page Pushing after Successful Authentication

The automatic page pushing function after the successful authentication means that the web page which user needs to access can be re-opened after the authentication. According to the actual situation, the welcome page before the automatic pushing authentication or the appointed web page by the automatic pushing function can be configured.

6.11.5.2 Automatic Page Pushing after Successful Authentication Configuration

Automatic Page Pushing after Successful Authentication Configuration is as below:

1. Enable/disable the captive portal authentication function
2. Configure the automatic page pushing after successful authentication

1. Enable/disable the captive portal authentication function

Command	Explanation
Captive Portal Mode	
enable **disable**	Enable/disable the captive portal authentication function on the switch.

2. Configure the automatic page pushing after successful authentication

Command	Explanation
Captive Portal Instance Mode	
redirect attribute url-after-login enable **no redirect attribute url-after-login enable**	Enable the function that the redirect url carries the pushed url after the successful authentication. The no command disables it.
redirect attribute url-after-login name <name> **no redirect attribute url-after-login name**	Configure the attribute name of the pushed url after the successful authentication which is carried in the redirect url. The no command recovers the name to be the default value.
redirect attribute url-after-login encode {plain-text|base64}	Configure the encode of the pushed url after the successful authentication which is carried in the redirect url.
redirect attribute url-after-login value <url-value> **no redirect attribute url-after-login value**	Configure the appointed url which is popped up after the successful authentication. The no command deletes it.

6.11.5.3 Automatic Page Pushing after Successful Authentication Example

Case:

Configure the automatic page pushing after successful authentication on configuration1 of http://www.test.com.

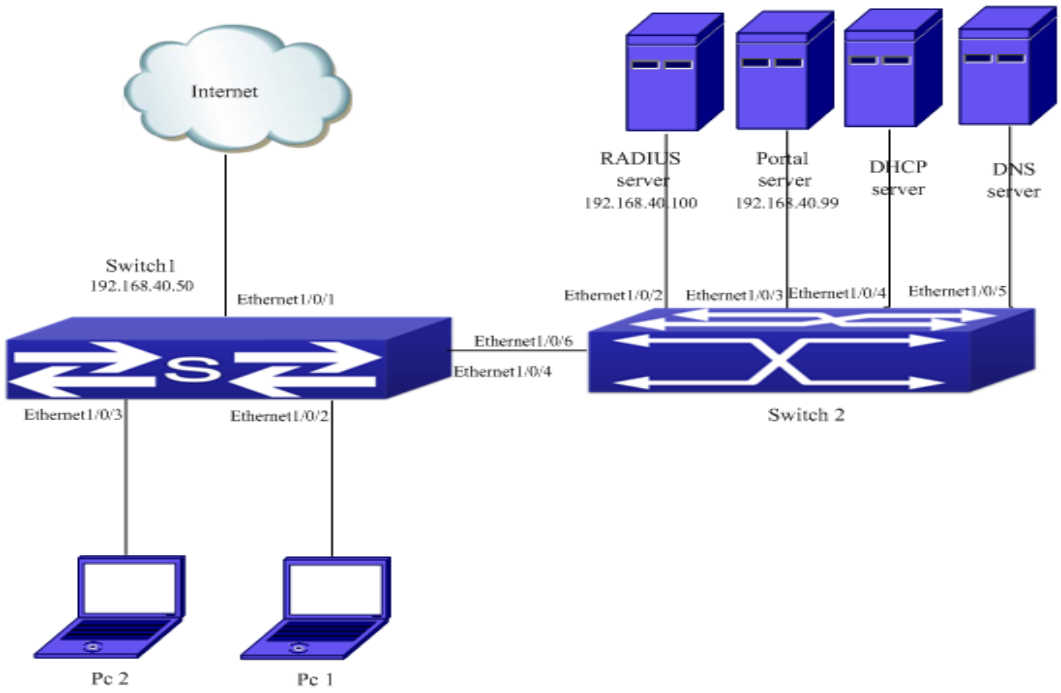


Figure 6-28 Automatic Page Pushing after Successful Authentication Configuration

Configuration steps:

Configure the portal server information for switch1.

switch1 (config-cp)#enable

switch1 (config-cp)#configuration 1

switch1 (config-cp-instance)#enable

switch1 (config-cp-instance)# redirect attribute url-after-login enable

switch1 (config-cp-instance)# redirect attribute url-after-login encode plain-text

switch1 (config-cp-instance)# redirect attribute url-after-login name ad

switch1 (config-cp-instance)# redirect attribute url-after-login value http://www.test.com

6.11.5.4 Automatic Page Pushing after Successful Authentication Troubleshooting

When there is problem in using the automatic page pushing function after the successful authentication, please check the following reasons:

- Check if the captive portal authentication function is configured correctly. The automatic page pushing function after the successful authentication can be effect when the captive portal function is normal.
- If the command of **redirect attribute url-after-login value** is configured, the configured page url can be pushed automatically after the authentication; if that command is not configured, the page that the user access before the authentication can be pushed.
- Check if the page before the authentication or the pushed page appointed by command exists, if not, the page cannot be accessed after pushing.

6.11.6 http-redirect-filter

6.11.6.1 Introduction to http-redirect-filter

http-redirect-filter appoints the IP or domain name for the HTTP redirection of portal authentication. Only the HTTP packet configured this IP or domain name can be redirected.

This function can be used with Captive Portal, which means to achieve filtering the user accessing after enabled captive portal.

6.11.6.2 http-redirect-filter Configuration

http-redirect-filter configuration task list:

1. Configure the http-redirect-filter rule
2. Bind the http-redirect-filter rule to cp instance

1. Configure the http-redirect-filter rule

Command	Explanation
Captive Portal Mode	
http-redirect-filter <1-32> (ip A.B.C.D| domain WORD) **no http-redirect-filter (<1-32>|all)**	Configure the http-redirect-filter rule. The no communicated deletes it.

2. Bind the http-redirect-filter rule to cp instance

Command	Explanation
Captive Portal Mode	
http-redirect-filter <1-32> **no http-redirect-filter (<1-32>|all)**	Bind a rule to a instance of the captive portal. The no command deletes the redirect binding.

6.11.6.3 http-redirect-filter Examples

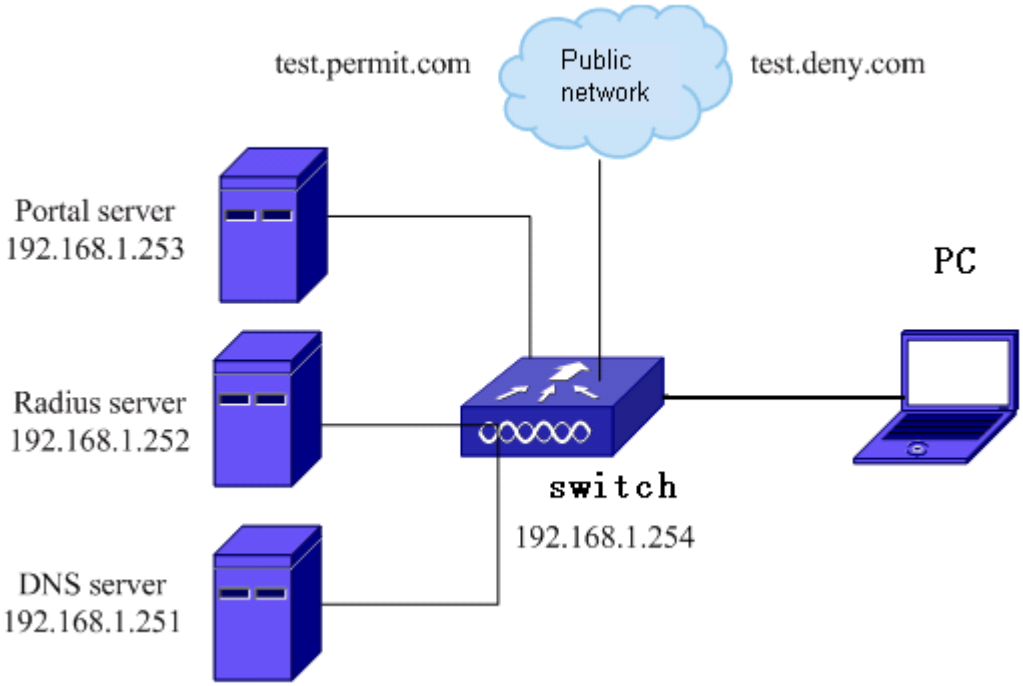


Figure 6-29 http-redirect-filter function case

As shown above, if client wants to access the public network of “test.permit.com” before the portal authentication, the url white list should be configured. If client wants to be forbidden to access the public network of “test.deny.com” after the portal authentication, the url black list should be configured.

Configure with the following steps:

1. Configure the related authentication key, authentication server, accounting server and aaa mode of the radius server under the global mode:

switch (config)#radius-server key 0 test

switch (config)#radius-server authentication host 192.168.1.252

switch (config)#radius-server accounting host 192.168.1.252

switch (config)#aaa-accounting enable

switch (config)#aaa enable

switch (config)#aaa group server radius radius_aaa_1

switch config-sg-radius)# server 192.168.1.252

switch(config)#interface vlan 192

switch(config-if-vlan1)#ip address 192.16.1.254 255.255.255.0

switch(config)#free-resource 1 destination ipv4 192.168.1.252/32

2. Configure the portal function, portal server under the portal instance:

Switch (config)#captive-portal

Switch (config-cp)#enable

Switch(config-cp)# nas-ipv4 192.168.1.254

Switch(config-cp)# external portal-server server-name abc ipv4 192.168.1.253

Switch (config-cp)# configuration 1
Switch (config-cp-instance)#enable
Switch (config-cp-instance)#name helix4
Switch (config-cp-instance)#radius accounting
Switch (config-cp-instance)#radius-acct-server abc99
Switch (config-cp-instance)#radius-auth-server abc99
Switch (config-cp-instance)#redirect attribute nas-ip enable
Switch (config-cp-instance)#ac-name helix4
Switch (config-cp-instance)#redirect url-head http://192.168.1.253/a70.htm
Switch (config-cp-instance)#portal-server ipv4 abc

3. Configure the http-redirect-filter rule:

Switch (config)#captive-portal
Switch (config-cp)# http-redirect-filter 1 domain test.permit.com
Switch (config-cp)#configuration 1
Switch (config-cp-instance)# http-redirect-filter 1

4. Enable the portal function on the port:

Switch (config)# interface ethernet1/0/2
Switch (config-if-ethernet1/0/2)#portal enable configuration 1

The client can be redirected authentication only through accessing “test.permit.com” before authentication. It cannot be redirected authentication by accessing other address.

6.11.6.4 http-redirect-filter Troubleshooting

If there are problems in using http-redirect-filter function, please check it with the following steps:

- ☞ Check if the configured rule is matching to the accessed domain name.
- ☞ If the DNS server configuration is correct and if it can analys the configured domain name correctly.
- ☞ If the command of http-redirect-filter under captive-portal configuration is configured.

6.11.7 Portal Non-perception

6.11.7.1 Introduction to Portal Non-perception

MAC authentication has the user experience that is "a authentication, multiple use". If you opened the fast authentication of MAC, the user successfully authenticates the first landing

Portal page, subsequent user can use any application.

In order to achieve a large number of user's fast authentication of MAC, user must use an external server to save the MAC binding information, and add it dynamically but not manually. This new realization of the program is called fast authentication of MAC scheme, since the user does not need to manually enter the user name and password for authentication when access network again, also known as Portal non-perception of authentication scheme.

6.11.7.2 Portal Non-perception Configuration

1. Enable/disable the quick mac authentication function

Command	Explanation
Captive Portal Config Mode	
fast-mac-auth **no fast-mac-auth**	Enable/disable the quick mac authentication function.

6.11.7.3 Portal Non-perception Examples

The created environment is as the following figure including the parts as below:

1. PC, user can access the network through the switch.

2. Public network, this part can be free or other switch devices.

3. Server, it includes:

MAC binding server, it is used to save the authenticated terminal mac address;

Radius server, it is used for the portal authentication and accounting;

Portal server, it is used for the portal authentication;

MAC binding server, Radius server and portal server can be the same one device. The mac binding server is the spread on the radius server.

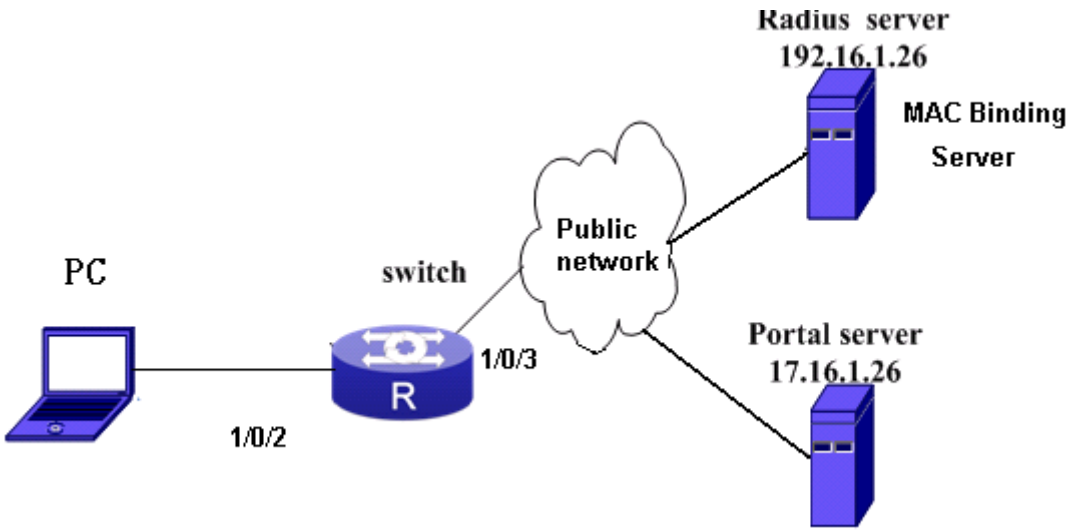


Figure 6-30

Configure as the following steps:

1. Configure the related authentication key, authentication server, accounting server and aaa mode of the radius server under the global mode:

switch (config)#radius-server key 0 test

switch (config)#radius-server authentication host 192.16.1.26

switch (config)#radius-server accounting host 192.16.1.26

switch (config)#aaa-accounting enable

switch (config)#aaa enable

switch (config)#aaa group server radius radius_aaa_1

switch config-sg-radius)# server 192.16.1.26

switch(config)#interface vlan 192

switch(config-if-vlan1)#ip address 192.16.1.50 255.255.255.0

switch(config)#free-resource 1 destination ipv4 192.16.1.26/32

2. Configure the portal function, portal server under the portal instance:

Switch (config)#captive-portal

Switch (config-cp)#enable

Switch(config-cp)# nas-ipv4 192.168.1.50

Switch(config-cp)# external portal-server server-name abc ipv4 172.16.1.26

Switch (config-cp)# configuration 1

Switch (config-cp-instance)#enable

Switch (config-cp-instance)#name helix4

Switch (config-cp-instance)#radius accounting

Switch (config-cp-instance)#radius-acct-server abc99

Switch (config-cp-instance)#radius-auth-server abc99

Switch (config-cp-instance)#redirect attribute nas-ip enable

Switch (config-cp-instance)#ac-name helix4

Switch (config-cp-instance)#redirect url-head http://172.16.1.26/a70.htm

Switch (config-cp-instance) # portal-server ipv4 abc

3. Configure the portal non-perception:

Switch (config-cp)#enable

Switch (config-cp)#fast-mac-auth

4. Enable the portal function on the port:

Switch (config)# interface ethernet1/0/2

Switch (config-if-ethernet1/0/2)#portal enable configuration 1

The normal portal authentication is needed in the first access. After the first time, user can use the non-perception authentication of portal.

6.11.7.4 Portal Non-perception Troubleshooting

Please check if the reasons are the following situations when there are problems in using the function of portal non-perception:

- ☞ Check whether the captive-portal function is enabled.
- ☞ Check whether the quick mac authentication function is enabled.
- ☞ Check whether issued the app table entry to the switch if the quick mac authentication is not effective after configured.

6.11.8 Portal Escaping

There is a risk in the current portal application. When the communication between the access device and portal server is broken, the new user cannot get on-line and the on-line user cannot get down; and the information of the access device and portal server is inconsistent. This will bring the accounting error. These phenomenons can bring the inconvenience to the operations and users.

The portal escaping function provides a good method to solve the above problems. It can make the user on-line and use the network normally when the portal server or radius server cannot working normally, and the new user can still access the network. So the portal escaping includes portal server escaping and radius server escaping.

6.11.8.1 Portal Server Escaping

6.11.8.1.1 Introduction to Portal Server Escaping

The principle of portal server escaping function is that: switch probes the portal server periodically. If the probing is successful, the server status will be configured as UP; if the probing failed N times (N can be configured), it will change the status of unreachable to be DOWN (escaping status), cancel the network authentication limit, allow the portal user accessing the network without authentication and send the trap and log information of the status changing. When it probes the server is reachable, it will recover the server status to be UP (authentication status), restart the network authentication limit, reject the user without authentication accessing the network and send the log and trap information of the status recovering.

The method that switch probes the portal server status is probing the TCP connection: switch launches the TCP connection to the portal server port of the portal server (the default is 2000, it can be configured) regularly. If the connection is successful, it means that this portal server is enabled, we consider that the probing is successful and the server is reachable (the status is UP); if the connection failed, we consider the probing failed.

Probing interval and maximum number of probing failures: the interval of the probing can

be configured through the command. The maximum number of probing failures means that the probing failures before that the server is reachable. One probing failure does not mean that the server is unreachable; user should view if the number of the probing failures achieves the configured maximum value. If the number achieves the configured value, the server can be considered as unreachable; otherwise, the number is just cumulative. After probing is successful, this number will be cleared to be 0. The probing interval and maximum number of probing failures can be configured through the command.

The server triggers the following three configurations when the status changes from reachable to unreachable, the administrator can select through the configuration:

- Send trap: send the trap information to the network management server. In the trap, it records the portal server name and the status information before and after the change of the server status.
- Send log: send the log information to the log server. In the log, it records the portal server name and the status information before and after the change of the server status.
- permit-all: it is also named as portal escaping. It means to cancel the portal authentication temporarily and allow all the portal users accessing the network when the portal server is in the unreachable status (down).

The server triggers the following three configurations when the status changes from unreachable to reachable. “Send trap” and “send log” can be selected through the configuration; “Disable portal escaping” is enforced to carry on:

- Send trap: send the trap information to the network management server. In the trap, it records the portal server name and the status information before and after the change of the server status.
- Send log: send the log information to the log server. In the log, it records the portal server name and the status information before and after the change of the server status.
- Disable portal escaping: If the portal server status changes to the reachable status (up), the portal authentication function of VAP will be recovered. The new user must pass the portal authentication for accessing the network.

Notice: The portal escaping function can only achieve that the new and old users are not affected when accessing the network currently. For the situation that user cannot get down the line normally, there are other methods. For example, the portal server is recovered to be UP, the access device will enforce the user to get down the line and it ensure that user can get down the line normally.

6.11.8.1.2 Portal Escaping Configuration

Portal escaping function configuration task list:

1. Enable the Portal escaping function and configure the probing interval and maximum number

of failures.

2. Show the current connection status of the Portal server.

1. Enable the Portal escaping function and configure the probing interval and maximum number of failures

Command	Explanation
Captive Portal Global Configuration Mode	
portal-server-detect server-name <name> [interval <interval>] [retry <retries>][action {log | permit-all | trap }] **no portal-server-detect server-name <name>**	Enable the Portal server escaping function and configure the related parameters (selectable) and the server configuration of status changing.

2. Show the current connection status of the Portal server

Command	Explanation
Admin Mode	
show captive-portal ext-portal-server server-name <name> status	Show the portal server status including the server address and if the portal escaping function is enabled.

6.11.8.1.3 Portal Escaping Examples

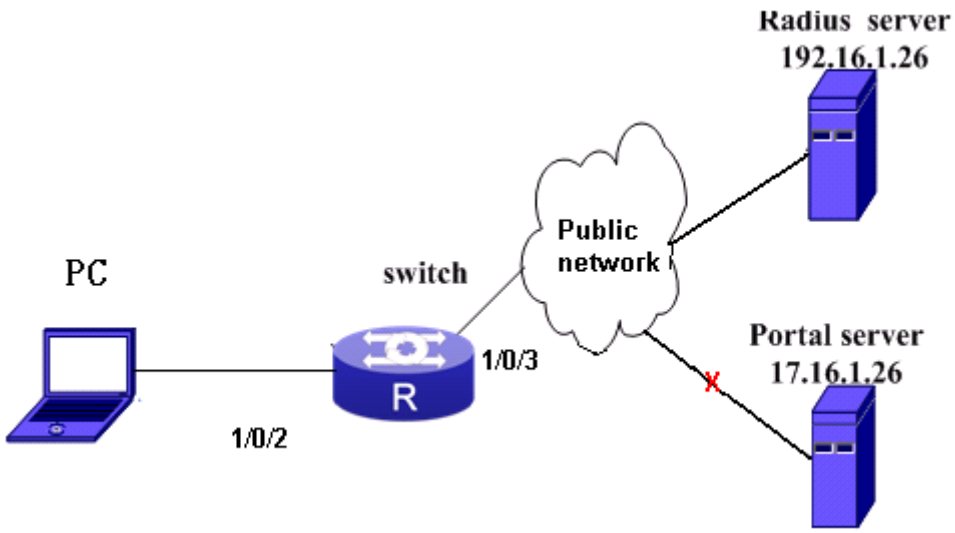


Figure 6-221 Portal server escaping function case

As shown above, in the situation of the normal working of portal server, the portal authentication can be normal for the network accessing when client is on-line. When the portal server is down or the connection between it and switch is broken, client cannot authenticate to on-line if the portal escaping function is not enabled on switch. If the portal escaping function is enabled on switch, switch can probe that the portal server is unavailable and start the portal escaping function. And the client can access the network without authentication. If client has passed the authentication before the portal server is broken, it will not be affected and it can still access the network.

The configuration is as below:

1. Configure the related authentication key, authentication server, accounting server and aaa mode of the RADIUS server in global mode.

```
switch (config)#radius-server key 0 test
switch (config)#radius-server authentication host 192.16.1.26
switch (config)#radius-server accounting host 192.16.1.26
switch (config)#aaa-accounting enable
switch (config)#aaa enable
switch (config)#aaa group server radius radius_aaa_1
switch config-sg-radius)# server 192.16.1.26
switch(config)#interface vlan 192
switch(config-if-vlan1)#ip address 192.16.1.50 255.255.255.0
switch(config)#free-resource 1 destination ipv4 192.16.1.26/32
```

2. Configure the portal function, portal server under the portal instance:

```
Switch (config)#captive-portal
Switch (config-cp)#enable
Switch(config-cp)# nas-ipv4 192.168.1.50
Switch(config-cp)# external portal-server server-name abc ipv4 172.16.1.26
Switch (config-cp)# configuration 1
Switch (config-cp-instance)#enable
Switch (config-cp-instance)#name helix4
Switch (config-cp-instance)#radius accounting
Switch (config-cp-instance)#radius-acct-server abc99
Switch (config-cp-instance)#radius-auth-server abc99
Switch (config-cp-instance)#redirect attribute nas-ip enable
Switch (config-cp-instance)#ac-name helix4
Switch (config-cp-instance)#redirect url-head http://172.16.1.26/a70.htm
Switch (config-cp-instance) # portal-server ipv4 abc
```

3. Configure the portal escaping:

```
Switch (config-cp)#enable
Switch (config-cp)# portal-server-detect server-name abc interval 600 retry 2 action log permit-all trap
```

4. Enable the portal function on the port:

```
Switch (config)# interface ethernet1/0/2
Switch (config-if-ethernet1/0/2)#portal enable configuration 1
```

As shown above, the portal server of cmcc is bound to CP instance and the probing function is configured; the probing interval is 600s. If the probing failed twice, send the trap information and log of the unreachable server and the enable the portal escaping function to allow the user without authentication accessing the network.

6.11.8.1.4 Portal Escaping Troubleshooting

In using, please adopt the following methods if the portal escaping function cannot be effective.

- ☞ Use the command of **show captive-portal ext-portal-server server-name <name> status** to check if the detect mode of the portal server is “enable”. If it is not “enable”, it means that the portal server escaping function is not enabled, please enable it.
- ☞ If the portal escaping function is enabled, check if the Detect Operational Status is down. Only when the server status is down, the portal escaping function can be enabled.
- ☞ If the portal server status is down, and the escaping function cannot be effective, the device may have the problem. Please contact to the sales engineers.

6.11.8.2 Radius Server Escaping

6.11.8.2.1 Introduction to Radius Server Escaping

After enabled the radius server escaping function, the escaping function is effective when all the radius servers cannot achieve, and the traffic of the portal authentication client will be allowed. When the authentication server is detected achieving again, the allowing rule will be deleted.

6.11.8.2.2 Radius Server Escaping Configuration

Radius server escaping configuration task list:
1. Enable the radius server escaping function
2. Configure the detection interval of radius server

1. Enable the radius server escaping function

Command	Explanation
Global Mode	
radius-server escape {enable | disable }	Enable the radius server escaping function.

2. Configure the detection interval of radius server

Command	Explanation
Global Mode	
radius-server escape detection-interval {default | <second>}	Configure the detection interval of radius server and the default value is 180s.

6.11.8.2.3 Radius Server Escaping Examples

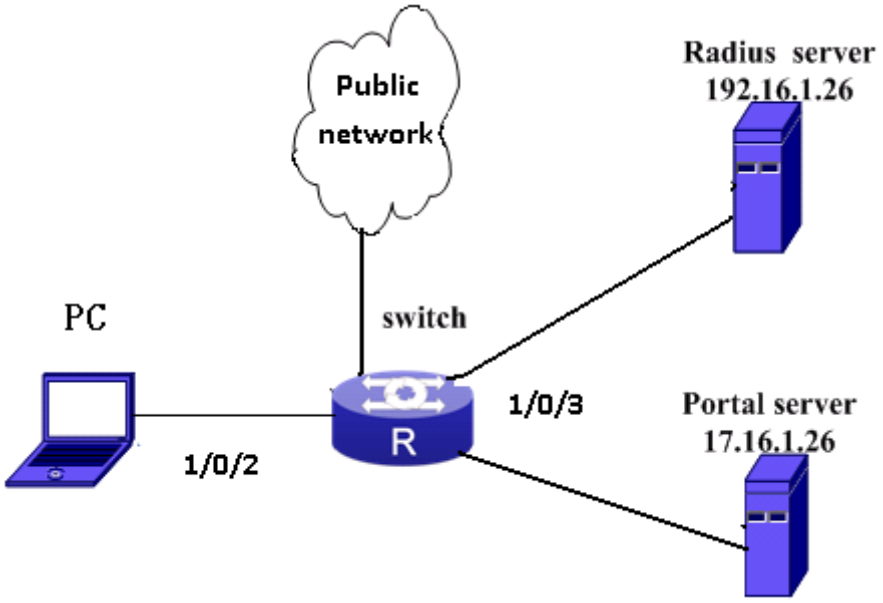


Figure 6-232 Radius server escaping case

The configuration is as below:

1 Configure the related authentication key, authentication server, accounting server and aaa mode of the RADIUS server in global mode.

switch (config)#radius-server key 0 test

switch (config)#radius-server authentication host 192.16.1.26

switch (config)#radius-server accounting host 192.16.1.26

switch (config)#aaa-accounting enable

switch (config)#aaa enable

switch (config)#aaa group server radius radius_aaa_1

switch config-sg-radius)# server 192.16.1.26

switch(config)#interface vlan 192

switch(config-if-vlan1)#ip address 192.16.1.50 255.255.255.0

switch(config)#free-resource 1 destination ipv4 192.16.1.26/32

2 Configure the portal function, portal server under the portal instance:

Switch (config)#captive-portal

Switch (config-cp)#enable

Switch(config-cp)# nas-ipv4 192.168.1.50

Switch(config-cp)# external portal-server server-name abc ipv4 172.16.1.26

Switch (config-cp)# configuration 1

Switch (config-cp-instance)#enable

Switch (config-cp-instance)#name helix4
Switch (config-cp-instance)#radius accounting
Switch (config-cp-instance)#radius-acct-server abc99
Switch (config-cp-instance)#radius-auth-server abc99
Switch (config-cp-instance)#redirect attribute nas-ip enable
Switch (config-cp-instance)#ac-name helix4
Switch (config-cp-instance)#redirect url-head http://172.16.1.26/a70.htm
Switch (config-cp-instance) # portal-server ipv4 abc

3 Configure the radius server escaping:
Switch (config)# radius-server escape enable
Switch (config)# radius-server escape detection-interval 120

4 Enable the portal function on the port:
Switch (config)# interface ethernet1/0/2
Switch (config-if-ethernet1/0/2)#portal enable configuration 1

The radius server is detected every 120s.

6.11.8.2.4 Radius Server Escaping Troubleshooting

If there are problems in using the radius server escaping function, please check it with the following steps:

- ☞ Check whether the radius server escaping function is enabled.
- ☞ If this function is enabled, get the parameters of Retransmit and Time Out by the command of **show aaa config** and then check whether if configured detection-interval is larger than the value of (Retransmit+1)*Time Out. If the detection-interval is smaller than it, the escaping function will fail.
- ☞ If the command of **show aaa config** shows Is Server live by radius escape function =0, it means the escaping is running but not effective. Please contact to the sales engineers.

6.12 QoS

6.12.1 QoS

6.12.1.1 Introduction to QoS

QoS (Quality of Service) is a set of capabilities that allow you to create differentiated services for network traffic, thereby providing better service for selected network traffic. QoS is a guarantee for service quality of consistent and predictable data transfer service to fulfill program requirements. QoS cannot generate extra bandwidth but provides more effective bandwidth management according to the application requirement and network management policy.

6.12.1.1.1 QoS Terms

QoS: Quality of Service, provides a guarantee for service quality of consistent and predictable data transfer service to fulfill program requirements. QoS cannot generate new bandwidth but provides more effective bandwidth management according to the application requirement and network management.

QoS Domain: QoS Domain supports QoS devices to form a net-topology that provides Quality of Service, so this topology is defined as QoS Domain.

CoS: Class of Service, the classification information carried by Layer 2 802.1Q frames, taking 3 bits of the Tag field in frame header, is called user priority level in the range of 0 to 7.

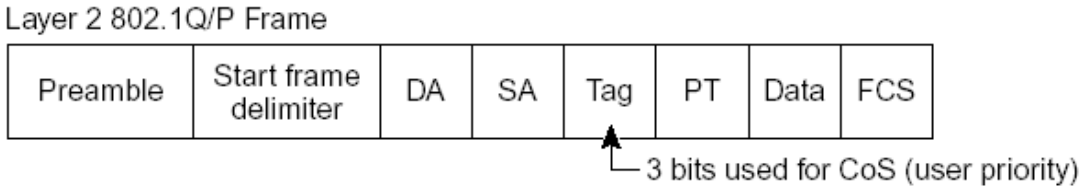


Figure 6-243 CoS priority

ToS: Type of Service, a one-byte field carried in Layer 3 IPv4 packet header to symbolize the service type of IP packets. Among ToS field can be IP Precedence value or DSCP value.

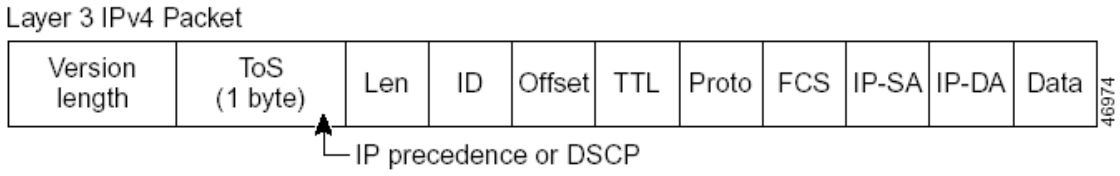


Figure 6-34 ToS priority

IP Precedence: IP priority. Classification information carried in Layer 3 IP packet header, occupying 3 bits, in the range of 0 to 7.

DSCP: Differentiated Services Code Point, classification information carried in Layer 3 IP packet header, occupying 6 bits, in the range of 0 to 63, and is downward compatible with IP Precedence.

MPLS TC*(EXP)*:

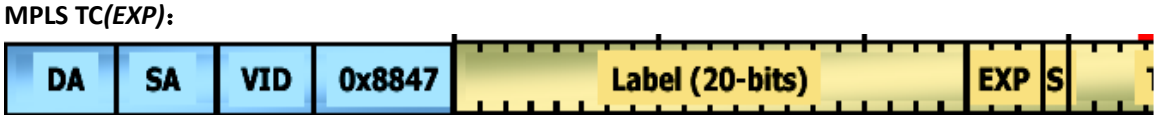


Figure 6-35

A field of the MPLS packets means the service class, there are 3 bits, the ranging from 0 to 7.

Internal Priority: The internal priority setting of the switch chip, it's valid range relates with the

chip, it's shortening is Int-Prio or IntP.

Drop Precedence: When processing the packets, firstly drop the packets with the bigger drop precedence, the ranging is the ranging is 0-2 in three color algorithm, the ranging is 0-1 in dual color algorithm. It's shortening is Drop-Prec or DP.

Classification: The entry action of QoS, classifying packet traffic according to the classification information carried in the packet and ACLs.

Policing: Ingress action of QoS that lays down the policing policy and manages the classified packets.

Remark: Ingress action of QoS, perform allowing, degrading or discarding operations to packets according to the policing policies.

Scheduling: QoS egress action. Configure the weight for eight egress queues WRR (Weighted Round Robin).

In-Profile: Traffic within the QoS policing policy range (bandwidth or burst value) is called In-Profile.

Out-of-Profile: Traffic out the QoS policing policy range (bandwidth or burst value) is called Out-of-Profile.

6.12.1.1.2 QoS Implementation

To implement the switch software QoS, a general, mature reference model should be given. QoS can not create new bandwidth, but can maximize the adjustment and configuration for the current bandwidth resource. Fully implemented QoS can achieve complete management over the network traffic. The following is as accurate as possible a description of QoS.

The data transfer specifications of IP cover only addresses and services of source and destination, and ensure correct packet transmission using OSI layer 4 or above protocols such as TCP. However, rather than provide a mechanism for providing and protecting packet transmission bandwidth, IP provide bandwidth service by the best effort. This is acceptable for services like Mail and FTP, but for increasing multimedia business data and e-business data transmission, this best effort method cannot satisfy the bandwidth and low-lag requirement.

Based on differentiated service, QoS specifies a priority for each packet at the ingress. The classification information is carried in Layer 3 IP packet header or Layer 2 802.1Q frame header. QoS provides same service to packets of the same priority, while offers different operations for packets of different priority. QoS-enabled switch or router can provide different bandwidth according to the packet classification information, and can remark on the classification information according to the policing policies configured, and may discard some low priority packets in case of bandwidth shortage.

If devices of each hop in a network support differentiated service, an end-to-end QoS solution can be created. QoS configuration is flexible, the complexity or simplicity depends on the network topology and devices and analysis to incoming/outgoing traffic.

6.12.1.1.3 Basic QoS Model

The basic QoS consists of four parts: Classification, Policing, Remark and Scheduling, where classification, policing and remark are sequential ingress actions, and Queuing and Scheduling are QoS egress actions.

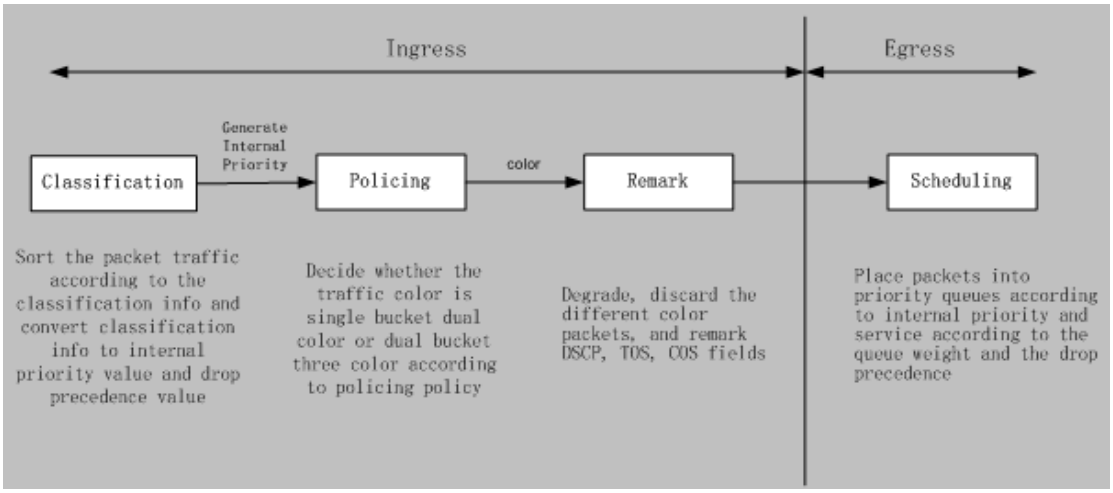


Figure 6-36 Basic QoS Model

Classification: Classify traffic according to packet classification information and generate internal priority and drop precedence based the classification information. For different packet types and switch configurations, classification is performed differently; the flowchart below explains this in detail.

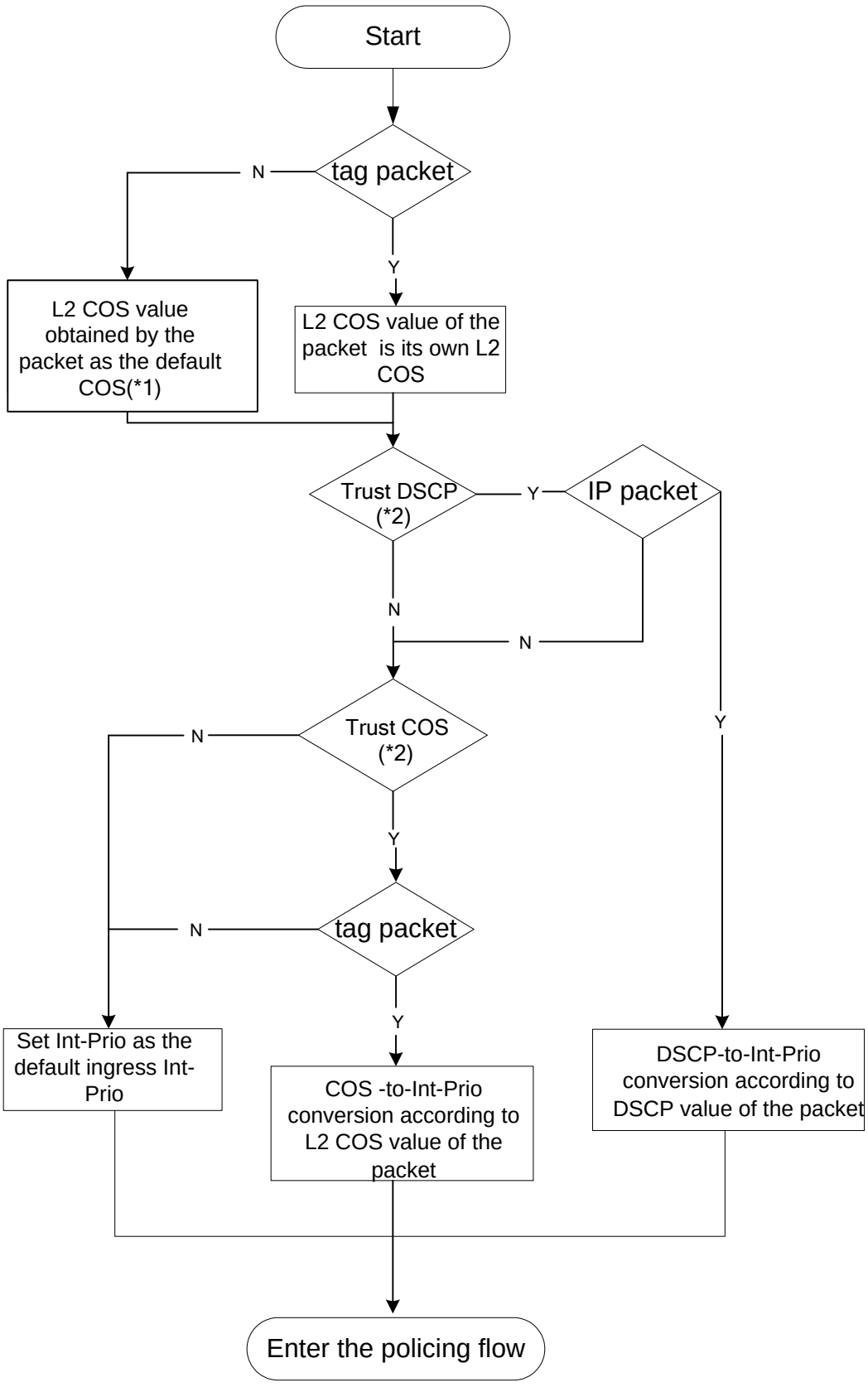


Figure 6-37 Classification process

Policing and remark: Each packet in classified ingress traffic is assigned an internal priority value and a drop precedence value, and can be policed and remarked.

Policing can be performed based on the flow to configure different policies that allocate bandwidth to classified traffic, the assigned bandwidth policy may be dual bucket dual color or dual bucket three color. The traffic, will be assigned with different color, can be discarded or passed, for the passed packets, add the remarking action. Remarking uses a new DSCP value of lower priority to replace the original higher level DSCP value in the packet. The following flowchart describes the operations.

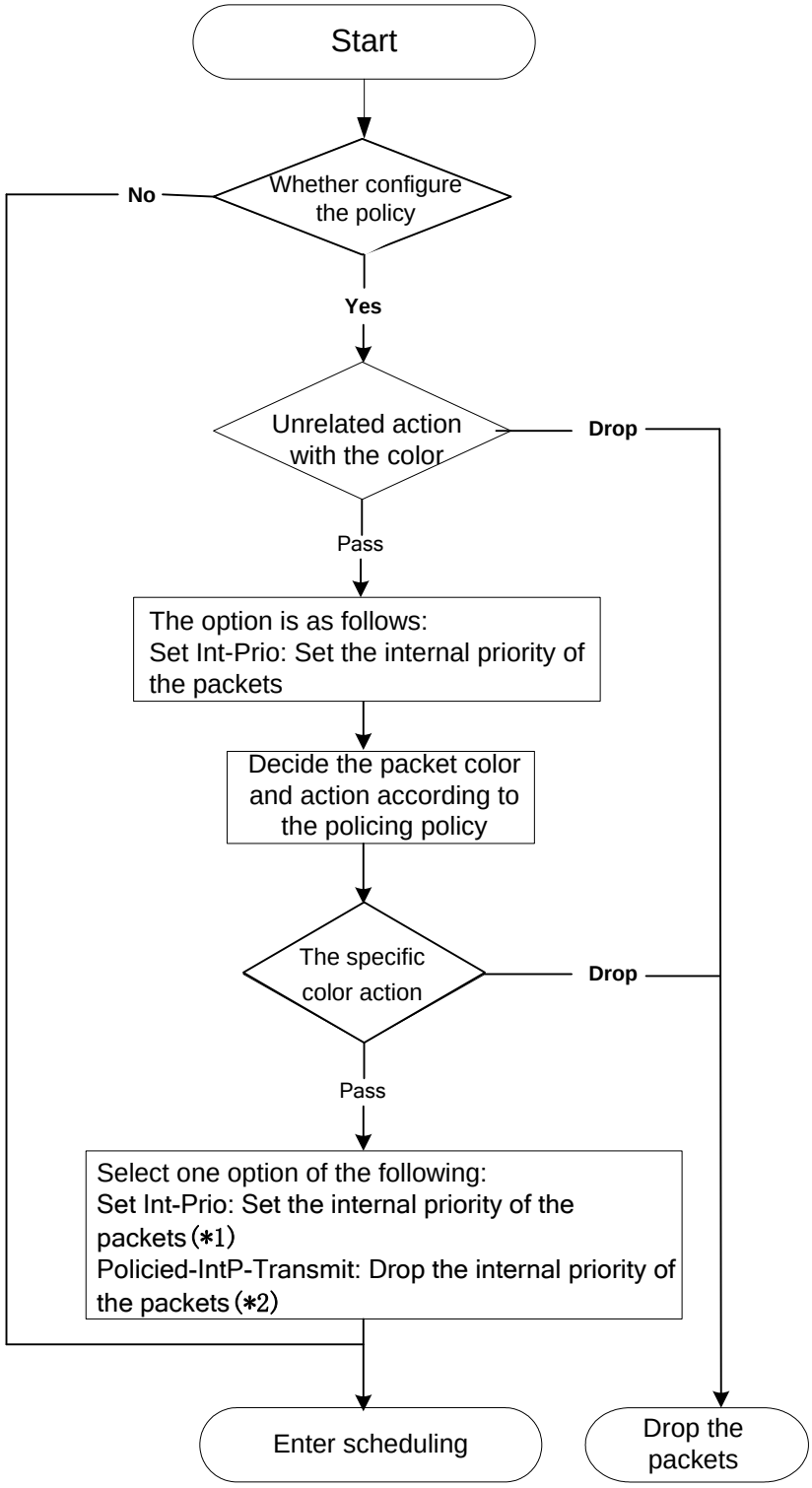


Figure 6-38 Policing and Remarking process

Queuing and scheduling: There are the internal priority and the drop precedence for the egress packets, the queuing operation assigns the packets to different priority queues according to the internal priority, while the scheduling operation perform the packet forwarding according to the priority queue weight and the drop precedence. The following flowchart describes the operations during queuing and scheduling.

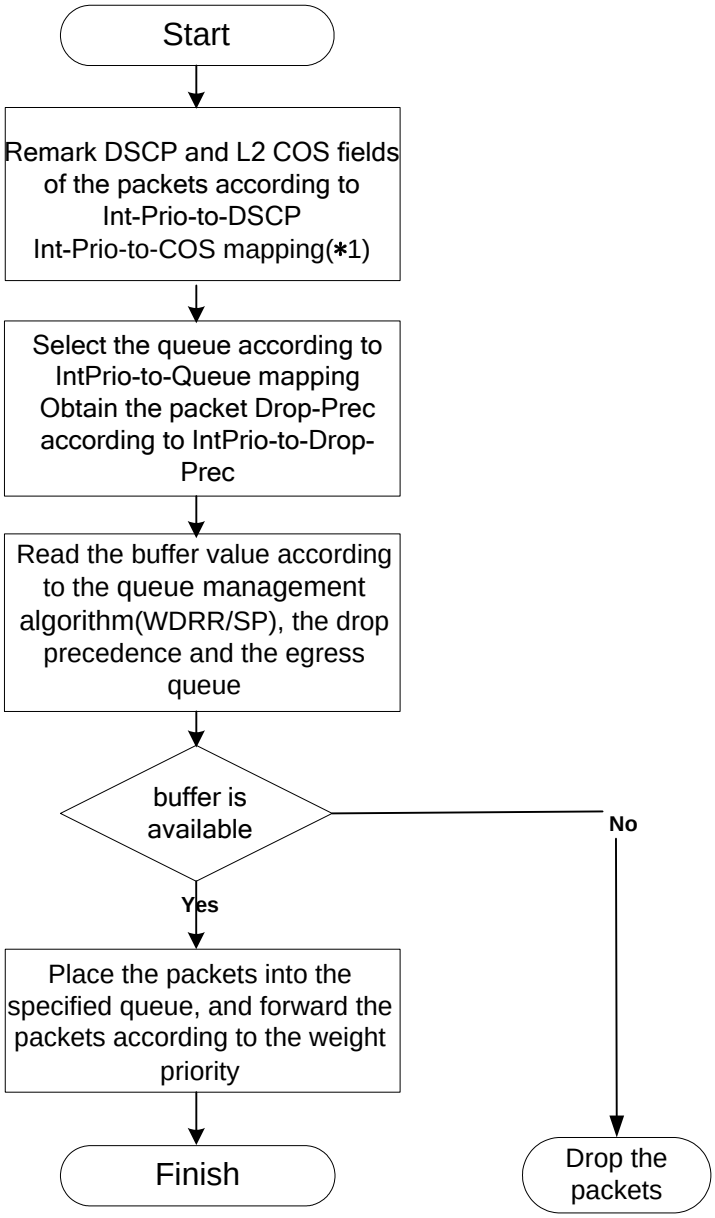


Figure 6-39 Queuing and Scheduling process

6.12.1.2 QoS Configuration Task List

Configure class map

Set up a classification rule according to ACL, CoS, VLAN ID, IPv4 Precedent, DSCP, IPV6 FL to classify the data stream. Different classes of data streams will be processed with different policies.

Configure a policy map

After data steam classification, a policy map can be created to associate with the class map created earlier and enter class mode. Then different policies (such as bandwidth limit, priority degrading assigning new DSCP value) can be applied to different data streams. You can also define a policy set that can be use in a policy map by several classes.

Apply QoS to the ports or the VLAN interfaces

Configure the trust mode for ports or bind policies to ports. A policy will only take effect on a port when it is bound to that port.

Apply QoS to the ports or the VLAN interfaces

Configure the trust mode for ports or bind policies to ports. A policy will only take effect on a port when it is bound to that port.

The policy may be bound to the specific VLAN, it also supports to modify dynamically.

It is not recommended to synchronously use policy map on VLAN and its port.

Configure queue management algorithm

Configure queue management algorithm, such as sp, wrr, wdrr, and so on.

Configure QoS mapping

Configure the mapping from CoS to DP, DSCP to DSCP, IntP or DP, IntP to DSCP.

1. Configure class map.

Command	Explanation
Global Mode	
class-map <*class-map-name*> **no class-map <class-map-name>**	Create a class map and enter class map mode; the "no **class-map <*class-map-name*>**" command deletes the specified class map.
match {access-group <acl-index-or-name> | ip dscp <dscp-list>| ip precedence <ip-precedence-list>| ipv6 access-group <acl-index-or-name> | ipv6 dscp <dscp-list>| ipv6 flowlabel <flowlabel-list>|vlan <vlan-list> / cos <*cos-list*> | exp <exp-list>} **no *match* {access-group | ip dscp | ip precedence / ipv6 access-group | ipv6 dscp | ipv6 flowlabel | vlan | cos | exp}**	Set matching criterion (classify data stream by ACL, CoS, VLAN ID, IPv4 Precedent, IPv6 FL or DSCP, etc) for the class map; the no command deletes specified matching criterion.

2. Configure a policy map

Command	Explanation
Global Mode	
policy-map <policy-map-name>	Create a policy map and enter policy map

no policy-map <policy-map-name>	mode; the no command deletes the specified policy map.
class <class-map-name> [insert-before <class-map-name>] **no class <class-map-name>**	After a policy map is created, it can be associated to a class. Different policy or new DSCP value can be applied to different data streams in class mode; the no command deletes the specified class.
set {ip dscp <*new-dscp*> | ip precedence <new-precedence> | internal priority <*new-inp*> | drop precedence <new-dp> | cos <new-cos>} **no set {ip dscp | *ip* precedence / *internal* priority | drop precedence | cos }K**	Assign a new DSCP, CoS, IP Precedence value for the classified traffic; the no command cancels the newly assigned value.
Single bucket mode: **policy <bits_per_second> <normal_burst_bytes> ({conform-action ACTION | exceed-action ACTION*}*)** **Dual bucket mode:** **policy <bits_per_second> <normal_burst_*bytes*> *[pir* <peak_rate_bps>] | <maximum_burst_bytes> [{conform-action ACTION | exceed-action ACTION | violate-action ACTION }]** **ACTION definition:** **drop | transmit | set-dscp-transmit <dscp_value> | set-prec-transmit <ip_precedence_value> | set-cos-transmit <*cos*_value> / set-internal-priority <inp_value> | set-Drop-Precedence <dp_value>** **no policy**	Configure a policy for the classified flow. The non-aggregation policy command supports three colors. Analyze the working mode of the token bucket, whether it is singe **rate** single bucket, single rate dual bucket, dual rate dual bucket, set corresponding action to different color packets. The no command will delete the mode configuration. Single bucket mode is supported by the specific switch.
policy aggregate <aggregate-policy-name> ***no*** **policy aggregate <aggregate-policy-name>**	Apply a policy to classified traffic; the no command deletes the specified policy set.
accounting **no accounting**	Set statistic function for the classified traffic. After enable this function under the policy class map mode, add statistic function to the traffic of the policy class map. In single bucket mode, the messages can only red or green when passing policy. In the print information, in-profile means green and out-profile means red. In dual bucket mode, there are three colors of the packets. In the print information, in-profile means green and out-profile means red and yellow.

Policy class map configuration *mode*	
drop **no drop** **transmit** **no transmit**	Drop or transmit data package that match the class, the no command cancels the assigned action.

3. Apply QoS to port or VLAN interface

Command	Explanation
Interface Configuration Mode	
mls qos *trust dscp* **no mls qos *trust* dscp**	Configure port trust; the no command disables the current trust status of the port.
mls qos *cos* {<default-cos>} **no mls qos cos**	Configure the default CoS value of the port; the no command restores the default setting.
service-policy input <policy-map-name> **no service-policy input {<policy-map-name>}**	Apply a policy map to the specified port; the no command deletes the specified policy map applied to the port. Egress policy map is not supported yet or deletes all the policy maps applied on the ingress direction of the port
Global Mode	
service-policy input <*policy-map-name*> vlan <*vlan-list*> **no service-policy input {<*policy-map-name*>} *vlan* <vlan-list>**	Apply a policy map to the specified VLAN interface; the no command deletes the specified policy map applied to the VLAN interface or deletes all the policy maps applied in the ingress direction of the vlan interface .

4. Configure queue management algorithm and weight

Command	Explanation
Port Configuration Mode	
mls **qos queue algorithm {sp | wrr | wdrr}** no mls **qos queue algorithm**	Set queue management algorithm, the default queue management algorithm is wrr.
mls **qos queue wrr weight <weight0..weight7>** no mls **qos queue wrr weight**	Set queue weight based a port, the default queue weight is 1 2 3 4 5 6 7 8.
mls **qos queue wdrr weight <weight0..weight7>** no mls **qos queue wdrr weight**	Set queue weight based a port, the default queue weight is 10 20 40 80 160 320 640 1280.
mls **qos queue <queue-id> bandwidth**	Set bandwidth guarantee based a port.

<minimum-bandwidth> <maximum-bandwidth> no mls **qos queue <queue-id> bandwidth**	

5. Configure QoS mapping

Command	Explanation
Global Mode	
mls **qos map (cos-dp <dp1…dp8> || cos-intp < in-cos list > | dscp-dscp <in-dscp list> to <out-dscp> |dscp-intp <in-dscp list> to <intp>| dscp-dp <in-dscp list> to <dp> | intp-exp <exp1…exp8>)** no mls **qos map (cos-dp | dscp-dscp | dscp-intp | dscp-dp)** mls **qos map intp-exp <exp1…exp8>** no **mls qos map intp-exp**	Set the priority mapping for QoS, the no command restores the default mapping value.

6. Clear accounting data of the specific ports or VLANs

Command	Explanation
Admin Mode	
clear mls qos statistics [in | out] {interface <interface-name> | vlan <vlan-id>}	Clear the in or out directions accounting data of the specified ports or VLAN Policy Map.

7. Show configuration of QoS

Command	Explanation
Admin Mode	
show mls **qos maps [cos-dp | cos-intp |dscp-dscp | dscp-intp | dscp-dp |intp-exp]**	Display the configuration of QoS mapping.
show class-map [<*class-map-name*>]	Display the classified map information of QoS.
show policy-map [<*policy-map-name*>]	Display the policy map information of QoS.
show mls qos *{interface [<interface-id>]* [policy | queuing] | vlan <vlan-id>}	Display QoS configuration information on a port.
show mls qos *in* {*interface <interface-name>* policy | vlan <vlan-id>}	Show the policy configuration information of the port or vlan of in direction.
show mls qos *interface <interface-id> wred* [queue <queue-id>] [dp<dp>]	Show the wred parameter corresponding to dp value of the appointed queue under the port.

Show mls qos *vlan*	Show the qos information of VLAN interface.
show mls qos *aggregate-policy* [*policy-name]*	Show the configuration information of aggregate-policy.

6.12.1.3 QoS Example

Example 1:

Enable QoS function, change the queue out weight of port ethernet 1/0/1 to 1:1:2:2:4:4:8:8, set the port in trust CoS mode without changing DSCP value, and set the default CoS value of the port to 5.

The configuration steps are listed below:

Switch#config
Switch(config)#interface ethernet 1/0/1
Switch(Config-If-Ethernet1/0/1)# mls qos queue wrr weight 1 1 2 2 4 4 8 8
Switch(Config-If-Ethernet1/0/1)#mls qos cos 5

Configuration result:

When QoS enabled in Global Mode, the egress queue bandwidth proportion of port ethernet1/0/1 is 1:1:2:2:4:4:8:8. When packets have CoS value coming in through port ethernet1/0/1, it will be map to the queue out according to the CoS value, CoS value 0 to 7 correspond to queue out 1, 2, 3, 4, 5, 6, 7, 8 respectively. If the incoming packet has no CoS value, it is default to 5 and will be put in queue6. All passing packets would not have their DSCP values changed.

Example 2:

In port ethernet1/0/2, set the bandwidth for packets from segment 192.168.1.0 to 10 Mb/s, with a burst value of 4 MB, all packets exceed this bandwidth setting will be dropped.

The configuration steps are listed below:

Switch#config
Switch(config)#access-list 1 permit 192.168.1.0 0.0.0.255
Switch(config)#class-map c1
Switch(Config-ClassMap-c1)#match access-group 1
Switch(Config-ClassMap-c1)#exit
Switch(config)#policy-map p1
Switch(Config-PolicyMap-p1)#class c1
Switch(Config-PolicyMap-p1-Class-c1)#policy 10000 4000 exceed-action drop
Switch(Config-PolicyMap-p1-Class-c1)#exit
Switch(Config-PolicyMap-p1)#exit
Switch(config)#interface ethernet 1/0/2
Switch(Config-If-Ethernet1/0/2)#service-policy input p1

Configuration result:

An ACL name 1 is set to matching segment 192.168.1.0. Enable QoS globally, create a class map named c1, matching ACL1 in class map; create another policy map named p1 and refer to c1 in p1, set appropriate policies to limit bandwidth and burst value. Apply this policy map on port ethernet1/0/2. After the above settings done, bandwidth for packets from segment 192.168.1.0 through port ethernet 1/0/2 is set to 10 Mb/s, with a burst value of 4 MB, all packets exceed this bandwidth setting in that segment will be dropped.

Example 3:

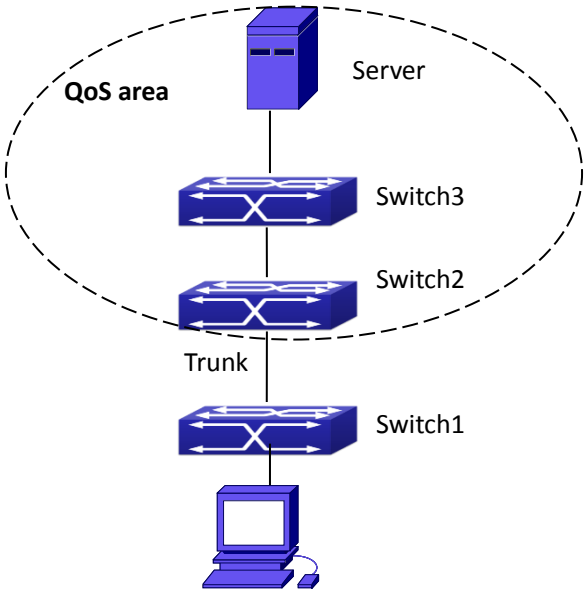


Figure 6-250 Typical QoS topology

As shown in the figure, inside the block is a QoS domain, Switch1 classifies different traffics and assigns different IP precedences. For example, set CoS precedence for packets from segment 192.168.1.0 to 5 on port ethernet1/0/1. The port connecting to switch2 is a trunk port. In Switch2, set port ethernet 1/0/1 that connecting to swtich1 to trust cos. Thus inside the QoS domain, packets of different priorities will go to different queues and get different bandwidth.

The configuration steps are listed below:

QoS configuration in Switch1:

Switch#config

Switch(config)#access-list 1 permit 192.168.1.0 0.0.0.255

Switch(config)#class-map c1

Switch(Config-ClassMap-c1)#match access-group 1

Switch(Config-ClassMap-c1)#exit

Switch(config)#policy-map p1

Switch(Config-PolicyMap-p1)#class c1

Switch(Config-PolicyMap-p1-Class-c1)#set ip precedence 5

```
Switch(Config-PolicyMap-p1-Class-c1)#exit
Switch(Config-PolicyMap-p1)#exit
Switch(config)#interface ethernet 1/0/1
Switch(Config-If-Ethernet1/0/1)#service-policy input p1
```

QoS configuration in Switch2:

```
Switch#config
Switch(config)#interface ethernet 1/0/1
```

6.12.1.4 QoS Troubleshooting

trust cos and exp can be used with other trust or Policy Map.

- ☞ trust dscp can be used with other trust or Policy Map. This configuration takes effect to IPv4 and IPv6 packets.
- ☞ trust exp, trust dscp and trust cos may be configured at the same time, the priority is: EXP>DSCP>COS.
- ☞ If the dynamic VLAN (mac vlan/voice vlan/ip subnet vlan/protocol vlan) is configured, then the packet COS value equals COS value of the dynamic VLAN.
- ☞ Policy map can only be bound to ingress direction, egress is not supported yet.
- ☞ At present, it is not recommended to synchronously use policy map on VLAN and VLAN's port.
- ☞

6.12.2 PBR

6.12.2.1 Introduction to PBR

PBR（Policy-Based Routing）is a method which determines the next-hop of the data packets by policy messages such as source address, destination address, IP priority, TOS value, IP protocol, source port No, destination port No, etc.

6.12.2.2 PBR Configuration

1. Configure a class-map
2. Set match standard of the class-map
3. Configure a policy-map
4. Configure a policy map corresponding to a class map
5. Configure nexthop IPv4 address

6. Configure the port binding policy map
7. Configure the VLAN binding policy map

1. Configure a class-map

Command	Explanation
Global Configuration Mode	
class-map <*class-map-name*> **no class-map <*class-map-name*>**	Set up or delete a class-map.

2. Set match standard of the class-map

Command	Explanation
Class-map Configuration Mode	
match ip {access-group <*acl-index-or-name*>} **no match ip {access-group}**	Set the match standard of the class-map

3. Configure a policy-map

Command	Explanation
Global Configuration Mode	
policy-map <*policy-map-name*> **no policy-map <*policy-map-name*>**	Set up or delete a policy-map.

4. Configure a policy map corresponding to a class map

Command	Explanation
Policy-map Configuration Mode	
class <*class-map-name*> **no class <*class-map-name*>**	Correspond a class-map, and enter the policy map mode.

5. Configure nexthop IPv4 address

Command	Explanation
Policy-class-map Mode	
set ipv4 [default] nexthop [vrf <vrf>] <*nexthop-ip*> **no set ipv4 nexthop**	Set nexthop IP for the classified traffic, the no command cancels the new assigned value.

6. Configure the port binding policy map

Command	Explanation
Port Mode	
service-policy {input <*policy-map-name*> | output <*policy-map-name*>} **no service-policy {input <*policy-map-name*> | output <*policy-map-name*>}**	Apply a policy map to the specified port. Only one policy map can be applied to each direction of each port. Egress policy map is not supported yet.

7. Configure the VLAN binding policy map

Command	Explanation
Global Configuration Mode	
service-policy input <*policy-map-name*> vlan <*vlan-list*> **no service-policy input <*policy-map-name*> vlan <*vlan-list*>**	Apply a policy map to the specified VLAN interface; the no command deletes the specified policy map applied to the VLAN interface.

6.12.2.3 PBR Examples

Example:

On port ethernet1/0/1, apply policy-based routing on packages from 192.168.1.0/24 segment, and set the next-hop as 218.31.1.119, meanwhile the local network IP of this network ranges within 192.168.0.0/16. To assure normal communication in local network, messages from 192.168.1.0/24 to local IP 192.168.0.0/16 are not applied with policy routing. The interface address of 192.168.1.0/24 of this device is 192.168.1.1.

Configuration procedure is as follows:

```
Switch#config
Switch(config)#ip access-list extended a1
Switch(Config-IP-Ext-Nacl-a1)# permit ip 192.168.1.0 0.0.0.255 192.168.0.0 0.0.255.255
Switch(Config-IP-Ext-Nacl-a1)#exit
Switch(config)#ip access-list extended a2
Switch(Config-IP-Ext-Nacl-a1)# permit ip 192.168.1.0 0.0.0.255 any-destination
Switch(Config-IP-Ext-Nacl-a1)#exit
Switch(config)#mls qos
Switch(config)#class-map c1
Switch(Config-ClassMap-c1)#match access-group a1
Switch(Config-ClassMap-c1)# exit
Switch(config)#class-map c2
Switch(Config-ClassMap-c2)#match access-group a2
Switch(Config-ClassMap-c2)# exit
Switch(config)#policy-map p1
Switch(Config-PolicyMap-p1)#class c1
Switch(Config-PolicyMap-p1-Class-c1)#set ip nexthop 192.168.1.1
Switch(Config-PolicyMap-p1-Class-c1)#exit
Switch(Config-PolicyMap-p1)#class c2
Switch(Config-PolicyMap-p1-Class-c2)#set ip nexthop 218.31.1.119
Switch(Config-PolicyMap-p1-Class-c2)#exit
Switch(config)#interface ethernet 1/0/1
Switch(Config-If-Ethernet1/0/1)#service-policy input p1
```

Configuration results:

First set ACL a1 and a2. a1 matches source IP segments 192.168.1.0/24 and destination IP segments 192.168.0.0/16. a2 matches source IP segments 192.168.1.0/24. Turn on QoS function in global mode and create two class-maps: c1 in which matches ACL a1 and c2 in which matches ACL a2. And create a policy-map in which quote c1. Set the interface address of 192.168.1.0/24 of this device as 192.168.1.1. Set the next-hop IP as 218.31.1.119 and apply the policy-map at port ethernet1/0/1. After that, all messages on port ethernet 1/0/1 from segment 192.168.1.0/24 will be transmitted through 218.31.1.119 except those from 192.168.0.0/16 segment which are still be transmitted through normal L3 routing.

6.12.3 IPv6 PBR

6.12.3.1 Introduction to PBR (Policy-based Router)

Policy-based routing provides a more powerful control over the forwarding and store of messages than traditional routing protocol to network managers. Traditionally, routers use the routing table derived from router protocol, and forward according to destination addresses. The policy-based router is more powerful and more flexible than the traditional one, because it enables network managers to choose the forwarding route not only according to destination addresses but also the size of messages, or source IP addresses. Policy can be defined as according to the balance of load in multiple routers or according to the quality of service (QOS) of the total flow forwarded in each line.

PBR (Policy-Based Routing) is a method which politically specifies the next hop when forwarding a data packet according to the source address, destination address, IP priority, TOS value, IP protocol, source port, destination port and other information of an IP packet.

6.12.3.2 PBR Configuration Task Sequence

1. Configure a class-map
2. Set the match standard in the class-map
3. Configure a policy-map
4. Configure to correlate a policy and a class-map
5. Configure the next hop IPv6 address
6. Configure the port binding policy map
7. Configure the VLAN binding policy map

1. Configure a class-map

Command	Explanation
Global Configuration Mode	
class-map <*class-map-name*> **no class-map <*class-map-name*>**	Create or delete a class-map.

2. Set the match standard in the class-map

Command	Explanation
Class-map Mode	
match ipv6 {access-group <*acl-index-or-name*>} **no match ipv6 {access-group }**	Set the match standard in the class-map.

3. Configure a policy-map

Command	Explanation
Global Configuration Mode	
policy-map <*policy-map-name*> **no policy-map <*policy-map-name*>**	Create or delete a policy-map.

4. Configure to correlate a policy and a class-map

Command	Explanation
Policy-map Mode	
class <*class-map-name*> **no class <*class-map-name*>**	Correlate with a class, and enter the policy-map mode.

5. Configure the next hop IPv6 address

Command	Explanation
Policy-class-map Mode	
set ipv6 [default] nexthop [vrf <vrf>] <*nexthop-ip*> **no set ipv6 nexthop**	Set the next hop IP for the classified flow, the no command cancels the new assigned value.

6. Configure the port binding policy-map

Command	Explanation
Port Configuration Mode	
service-policy {input <*policy-map-name*> | output <*policy-map-name*>} **no service-policy {input <*policy-map-name*> | output <*policy-map-name*>}**	Apply a policy map to the specified port. Only one policy map can be applied to each direction of each port. Egress policy map is not supported yet.

7. Configure the VLAN binding policy map

Command	Explanation
Global Mode	
service-policy input <*policy-map-name*> vlan <*vlan-list*> **no service-policy input <*policy-map-name*> vlan < *vlan-list* >**	Apply a policy map to the specified VLAN interface; the no command deletes the specified policy map applied to the VLAN interface.

6.12.3.3 PBR Examples

Example:

On port ethernet 1/0/1, the default gateway address of this device is 3000::1, set the messages whose source IP is within the segment 2000:: /64 to do policy routing, the next hop is 3100::2.

The following is the configuration steps:

```
Switch#config
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ipv6 address 2000::1/64
Switch(Config-if-Vlan1)#ipv6 neighbor 2000::2 00-00-00-00-00-01 interface Ethernet 1/0/1
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)#ipv6 address 3000::1/64
Switch(Config-if-Vlan2)#ipv6 neighbor 3000::2 00-00-00-00-00-02 interface Ethernet 1/0/2
Switch(config)#interface vlan 3
Switch(Config-if-Vlan3)#ipv6 address 3100::1/64
Switch(Config-if-Vlan3)#ipv6 neighbor 3100::2 00-00-00-00-00-03 interface Ethernet 1/0/5
Switch(config)# ipv6 access-list extended b1
Switch(Config-IPv6-Ext-Nacl-b1)# permit 2000:: /64 any-destination
Switch(Config-IPv6-Ext-Nacl-b1)#exit
Switch(config)#class-map c1
Switch(config-ClassMap)#match ipv6 access-group b1
Switch(config-ClassMap)# exit
Switch(config)#policy-map p1
Switch(config-PolicyMap)#class c1
Switch(config-Policy-Class)# set ipv6 nexthop 3100::2
Switch(config--Policy-Class)#exit
Switch(config-PolicyMap)#exit
Switch(config)#interface ethernet 1/0/1
Switch(Config-Ethernet1/0/1)#service-policy input p1
```

Configuration result:

First, set an ACL containing one entry, names it as b1, matching source IP segment 2000::/64(permit). Globally enable QoS function, create a class-map:c1, and match ACL b1 in the class-map. Create a policy-map:p1, quoting c1 in p1, and set the next hop as 3100::2. Apply this policy-map on port ethernet 1/0/1. After that, the messages whose source IP are within the segment 2000::/64 received on port ethernet 1/0/1 will be forwarded through 3100::2.

6.12.3.4 PBR Troubleshooting Help

- ☞ At present, policy-map can only be bound to input port but not output port.
- ☞ Since hardware resources are limited, if the policy is too complicated to configure, relative information will be noticed to users.

6.13 Flow-based Redirection

6.13.1 Introduction to Flow-based Redirection

Flow-based redirection function enables the switch to transmit the data frames meeting some special condition (specified by ACL) to another specified port. The fames meeting a same special condition are called a class of flow, the ingress port of the data frame is called the source port of redirection, and the specified egress port is called the destination port of redirection. Usually there are two kinds of application of flow-based redirection: 1. connecting a protocol analyzer (for example, Sniffer) or a RMON monitor to the destination port of redirection, to monitor and manage the network, and diagnose the problems in the network; 2. Special transmission policy for a special type of data frames.

The switch can only designate a single destination port of redirection for a same class of flow within a source port of redirection, while it can designate different destination ports of redirection for different classes of flows within a source port of redirection. The same class of flow can be applied to different source ports.

6.13.2 Flow-based Redirection Configuration Task Sequence

1.Flow-based redirection configuration
2.Check the current flow-based redirection configuration

1. Flow-based redirection configuration

Command	Explanation
Physical Interface Configuration Mode	
access-group <*aclname*> redirect to interface [ethernet <*IFNAME*>|<*IFNAME*>] **no access-group <*aclname*> redirect**	Specify flow-based redirection for the port; the "**no access-group <*aclname*> redirect**" command is used to delete flow-based redirection.

2. Check the current flow-based redirection configuration

Command	Explanation
Global Mode/Admin Mode	
show flow-based-redirect {interface [ethernet <*IFNAME*> |<*IFNAME*>]}	Display the information of current flow-based redirection in the system/port.

6.13.3 Flow-based Redirection Examples

Example:

User's request of configuration is listed as follows: redirecting the frames whose source IP is 192.168.1.111 received from port 1 to port 6, that is sending the frames whose source IP is 192.168.1.111 received from port 1 through port6.

Modification of configuration:

1: Set an ACL, the condition to be matched is: source IP is 192.168.1.111;

2: Apply the redirection based on this flow to port 1.

The following is the configuration procedure:

Switch(config)#access-list 1 permit host 192.168.1.111

Switch(config)#interface ethernet 1/0/1

Switch(Config-If-Ethernet1/0/1)# access-group 1 redirect to interface ethernet 1/0/6

6.13.4 Flow-based Redirection Troubleshooting Help

When the configuration of flow-based redirection fails, please check that whether it is the following reasons causing the problem:

The type of flow (ACL) can only be digital standard IP ACL, digital extensive IP ACL, nomenclature standard IP ACL, nomenclature extensive IP ACL, digital standard IPv6 ACL, and nomenclature standard IPv6 ACL;

☞Parameters of Timerange and Portrange can not be set in ACL, the type of ACL should be Permit.

☞The redirection **port** must **be** 1000Mb port in the flow-based redirection function.

☞Do not implement the forward across VLAN for flow-based redirection.

6.14 Egress QoS

6.14.1 Introduction to Egress QoS

In traditional IP networks, all packets are treated in the same way. All network equipments treat them by the first-in-first-out policy and try best effort to send them to the destination. However, it does not guarantee the performance like reliability and transmission delay. Network develops so fast that new demand has been raised for the quality of service on IP network with the continual emergence of new applications. For example, delay-sensitive services like VoIP and

video put higher demands on packet transmission delay and users cannot accept too long transmission delay (by contrast, E-mail and FTP services are not sensitive to the time delay). In order to support services with different service requirement like voice, video and data service, the network is required to be able to distinguish between different communications and provide appropriate service. The traditional best-effort IP network cannot identify and distinguish various kinds of communications while this ability is the very premise of providing differentiated services for different communications. Therefore, the best-effort service mode of traditional network cannot meet the demand of applications. The emergence of QoS techniques is committed to solve this problem.

Egress PolicyMap is the QoS policy in egress which performs QoS control of packets in the egress direction and provides better service for specified network communication with kinds of techniques. Egress PolicyMap includes class-map and policy-map, of which class-map is used for selecting packets to operate and policy-map is used for specifying the operation to use. Not all equipments support Egress QoS currently.

6.14.1.1 Egress QOS Terms

Egress QoS: Achieving QoS on egress of port.
Inner_vid: VLAN ID brought by the TAG near the header of network layer when double TAGs exist.
Outer_vid: VLAN ID brought by the TAG near the header of network link layer when double TAGs exist. The TAG is considered to be outer tag by default when only one TAG exists.
Outer_tpid: Protocol type of the network link layer header indicating the type of outer tag.

6.14.1.2 Basic Egress QoS Model

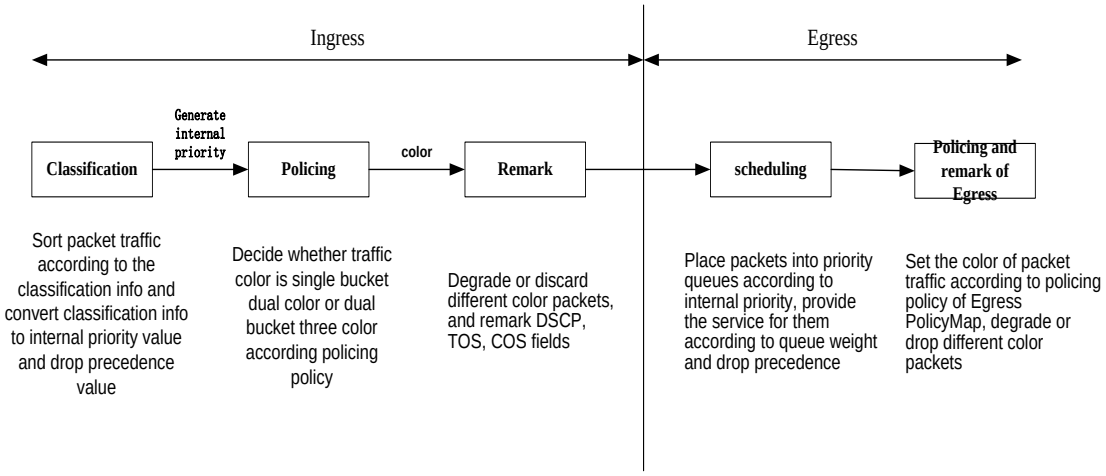


Figure 6-41

According to the characters (including field values like COS and DSCP) of upstream packets, policing and rewriting of Egress make the last QoS change on the packet prior to the packet egress.

Policing configures different policing policy based on the flow and distributes bandwidth for the flow classified. The distribution policy of bandwidth can be either dual bucket dual color or

dual bucket three color. Different colors can be assigned to different flows and approaches of discard or passage can be chosen for them; you can add rewriting action for packets with passage approach chosen. See the following flow chart for detailed description of Egress QoS:

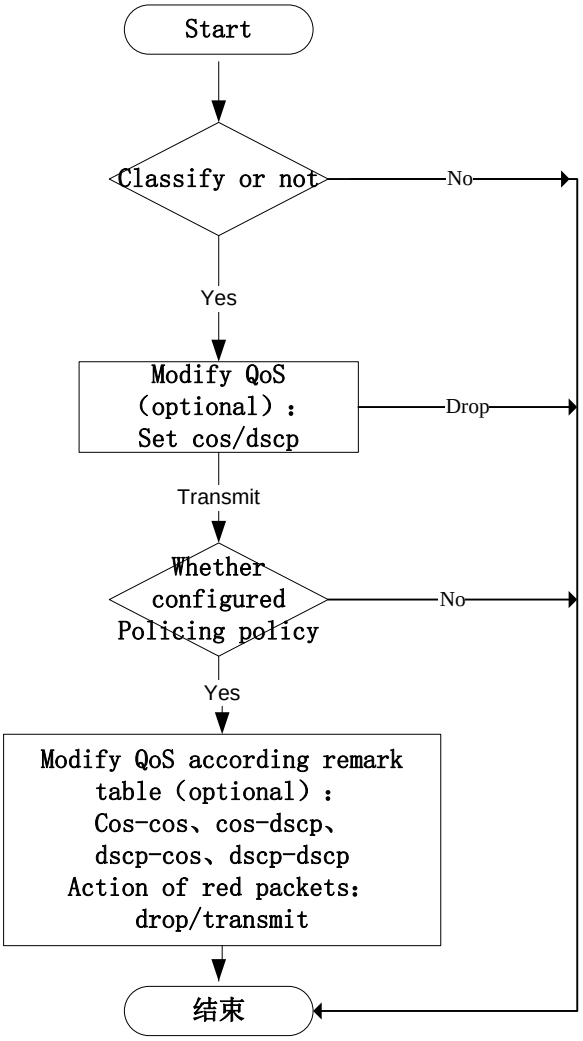


Figure 6-42

6.14.2 Egress QoS Configuration

Egress QoS Configuration Task List:

Configure class map

Set up a classification rule according to ACL, CoS, VLAN ID, IPv4 Precedent, DSCP, IPV6 DSCP to classify the data stream. Different classes of data streams will be processed with different policies.

Configure policy map

After data steam classification, a policy map can be created to associate with a class map created earlier and enter policy class mode. Then different policies (such as bandwidth limit, assigning new DSCP value) can be applied to different data streams.

Apply Egress QoS to port or VLAN

Configure the trust mode or binding policies for ports. A policy will only take effect on a port when it is bound to that port.

The policy may be bound to the specific VLAN, it also supports to modify dynamically.

1. Configure a class-map

Command	Explanation
Global Mode	
class-map <*class-map-name*> **no class-map <*class-map-name*>**	Create a class-map and enter class-map mode, no command deletes the specified class-map.
match {access-group <*acl-index-or-name*> | ip dscp <*dscp-list*> | ip precedence <*ip-precedence-list*> | ipv6 dscp <*dscp-list*> | vlan <*vlan-list*> | cos <*cos-list*> | ipv6 access-group <*acl-index-or-name*>} **no match {access-group | ip dscp | ip precedence | ipv6 dscp | vlan | cos | ipv6 access-group}**	Configure the matched standard of the class map to classify the data stream according to ACL, CoS, VLAN ID, IPv4 Precedence, DSCP, IPv6 DSCP priority; no command deletes the specific matched standard.

2. Configure a policy-map

Command	Explanation
Global Mode	
policy-map <*policy-map-name*> **no policy-map <*policy-map-name*>**	Create a policy-map and enter policy-map mode, no command deletes the specific policy-map.
class <*class-map-name*> [insert-before <*class-map-name*>] **no class <*class-map-name*>**	Create a policy map to associate with a class map and enter policy class map mode, then different data streams can apply different policies and be assigned a new DSCP value. No command deletes the specified policy class map.
set {ip dscp <*new-dscp*> | ip precedence <*new-precedence*> | cos <*new-cos*> | c-vid <*new-c-vid*> | s-vid <*new-s-vid*> | s-tpid <*new-s-tpid*>} **no set {ip dscp | ip precedence | cos | c-vid | s-vid | s-tpid}**	Assign a new DSCP, CoS and IP Precedence value for the classified flow, no command cancels the operation.
Single bucket mode: **policy <*bits_per_second*> <*normal_burst_bytes*> ({conform-action ACTION} | exceed-action ACTION})** **Dual bucket mode:** **policy <*bits_per_second*> <*normal_burst_bytes*> [pir <*peak_rate_bps*>] |**	Configure a policy for the classified flow. The non-aggregation policy command supports three colors. Analyze the working mode of the token bucket, whether it is single rate single bucket, single rate dual bucket or dual rate dual bucket, set corresponding action to different color packets. The no command will delete the configuration. Only specific switch supports

<*maximum_burst_bytes*> [{conform-action ACTION | exceed-action ACTION | violate-action ACTION}] **ACTION definition:** **drop | transmit | set-dscp-transmit <*dscp_value*> | set-cos-transmit <*cos_value*>** **no policy**	single bucket mode.
accounting **no accounting**	Set statistic function for the classified flow. After enable this function under the policy class map mode, add statistic function to the flow of the policy class map. In single bucket mode, packets can only red or green when passing policy. In the print information, in-profile means green and out-profile means red. In dual bucket mode, there are three colors of packets in-profile means green and out-profile means red and yellow.

3. Apply policy to port or VLAN

Command	Explanation
Interface Mode	
service-policy output <*policy-map-name*> **no service-policy output {<*policy-map-name*>}**	Apply a policy map to the egress of the port; the no command deletes the specified policy map applied to the port or deletes all the policy maps applied on the egress direction of the port .
Global Mode	
service-policy output <*policy-map-name*> vlan <*vlan-list*> **no service-policy output {<*policy-map-name*>} vlan <*vlan-list*>**	Apply a policy map to the egress of the VLAN; the no command deletes the specified policy map applied to the VLAN interfaceor deletes all the policy maps applied in the egress direction of the vlan interface .

4. Clear accounting data of the specific ports or VLANs

Command	Explanation
Admin Mode	
clear mls qos statistics out [interface <interface-name> | vlan <vlan-id>]	Clear the out direction accounting data of the specified ports or VLAN Policy Map.

5. Show QoS configuration

Command	Explanation
Admin Mode	
show mls qos {interface [<*interface-id*>] [policy | queuing] | vlan <*vlan-id*>}	Show QoS configuration of the port.
show class-map [<class-map-name>]	Show the class map information of QoS.
show policy-map [<policy-map-name>]	Show the policy map information of QoS.
show mls qos maps [cos-dp| cos-intp | dscp-dscp | dscp-intp | dscp-dp |intp-exp] | [begin | include | exclude <regular-expression>]	Show the configuration of QoS global mapping.

6.14.3 Egress QoS Examples

Example1:

On the egress of the port1, change cos value as 4 for the packet with dscp value of 0.

Create a class map:

switch(config)#class-map 1

switch(config-classmap-1)#match ip dscp 0

switch(config-classmap-1)#exit

Create a policy map:

switch(config)#policy-map 1

switch(config-policymap-1)#class 1

switch(config-policymap-1-class-1)#set cos 4

switch(config-policymap-1-class-1)#exit

switch(config-policymap-1)#exit

Bind a policy to the port:

switch(config)#in e 1/0/1

switch(config-if-ethernet1/0/1)#service-policy output 1

Example2:

On the egress of vlan10, change cos value as 4 for the packet with ipv6 dscp value of 7.

Create a class map:

switch(config)#class-map 1

switch(config-classmap-1)#match ipv6 dscp 7

switch(config-classmap-1)#exit

Create a policy map:

switch(config)#policy-map 1

```
switch(config-policymap-1)#class 1
switch(config-policymap-1-class-1)#set cos 4
switch(config-policymap-1-class-1)#exit
switch(config-policymap-1)#exit

Bind a policy to VLAN
switch(config)#service-policy output 1 vlan 10
```

6.14.4 Egress QoS Troubleshooting Help

- ☞ Not all equipments support Egress QoS presently, so please make sure the current device supports this function.
- ☞ If the policy configured cannot bind to the port or VLAN, please check whether the match option in classification table is supported by the current device.
- ☞ If terminal printing suggests lack of resource, please make sure there is enough resource to send the current policy.
- ☞ If the policy with match acl configured cannot bind to the port or VLAN, please make sure rules including permit exist in ACL.

6.15 Flexible QinQ

6.15.1 Introduction to Flexible QinQ

6.15.1.1 QinQ Technique

Dot1q-tunnel is also called QinQ (802.1Q-in-802.1Q), which is an expansion of 802.1Q. Its dominating idea is encapsulating the customer VLAN tag (CVLAN tag) to the service provider VLAN tag (SPVLAN tag). The packet with two VLAN tags is transmitted through the backbone network of the ISP internet to provide a simple layer-2 tunnel for the users. It is simple and easy to manage, applicable only by static configuration, and especially adaptive to small office network or small metropolitan area network using layer-3 switch as backbone equipment.

There are two kinds of QinQ: basic QinQ and flexible QinQ, the priority of flexible QinQ is higher than basic QinQ.

6.15.1.2 Basic QinQ

Basic QinQ based the port. After a port configures QinQ, whether the received packet with

tag or not, the device still packs the default VLAN tag for the packet. Using basic QinQ is simple, but the setting method of VLAN tag is inflexible.

6.15.1.3 Flexible QinQ

Flexible QinQ based data flow. It selects whether pack the external tag and packs what kind of the external tag by matching the material flow. For example: implement the property of flexible QinQ according to the user's VLAN tag, MAC address, IPv4/IPv6 address, IPv4/IPv6 protocol and the port ID of the application, etc. So, it can encapsulate the external tag for the packet and implements different scheme by different users or methods.

6.15.2 Flexible QinQ Configuration Task List

The match of flexible QinQ data flow uses policy-map rule of QoS to be sent, the configuration task list is as follows:

1. Create class-map to classify different data flows
2. Create flexible QinQ policy-map to relate with the class-map and set the corresponding operation
3. Bind flexible QinQ policy-map to port, it also supports to modify dynamically

1. Configure class map

Command	Explanation
Global mode	
class-map <*class-map-name*> **no class-map <*class-map-name*>**	Create a class-map and enter class-map mode, the no command deletes the specified class-map.
match {access-group <*acl-index-or-name*>|c-vlan <vlan-list>| ip dscp <dscp-list>| ip precedence <ip-precedence-list>| ipv6 access-group <acl-index-or-name>| ipv6 dscp <dscp-list> | ipv6 flowlabel <*flowlabel-list*> | vlan <*vlan-list*> | cos <*cos-list*>} **no match {access-group | ip dscp | ip precedence|ipv6 access-group| ipv6 dscp | ipv6 flowlabel | vlan | cos}**	Set the match standard of class-map, (classify data flow by ACL, CoS, VLAN ID,CVid, IPv4 Precedent or DSCP, etc for the class map); the no command deletes the specified match standard.

2. Configure policy-map of flexible QinQ

Command	Explanation
Global mode	
policy-map <*policy-map-name*> **no policy-map <*policy-map-name*>**	Create a policy-map and enter policy-map mode, the no command deletes the specified policy-map.
class <*class-map-name*> [insert-before	After a policy-map is created, it can be

<class-map-name>] **no class <*class-map-name*>**	associated to a class. Different policy or new DSCP value can be applied to different data flows in class mode; the no command deletes the specified class-map.
set {s-vid <new-vid> | s-tpid <0x8100 | 0x88a8 | 0x9100> | c-vid <new-vid >} **no set {c-vid | s-vid | s-tpid}**	Assign the new cos and vid value to the packets which match the class map, no command cancels the operation.
add {s-vid <new-vid> | c-vid <new -vid>} **no add {s-vid | c-vid}**	Assign the new SVid and CVid value to the packets which match the class map, no command cancels the operation.
delete c-vid **no delete c-vid**	Delete the inner VLAN Tag for the packet which match the class map, no command cancels the operation.

3. Bind flexible QinQ policy-map to port

Command	Explanation
Port mode	
service-policy {input < policy-map-name >} **no service-policy {input *[*policy-map-name]}**	Apply a policy-map to a port, the no command deletes the specified policy-map applied to the port.

4. Show flexible QinQ policy-map bound to port

Command	Explanation
Admin mode	
show mls *qos* {interface [<interface-id>]	Show flexible QinQ configuration on the port.

6.15.3 Flexible QinQ Example

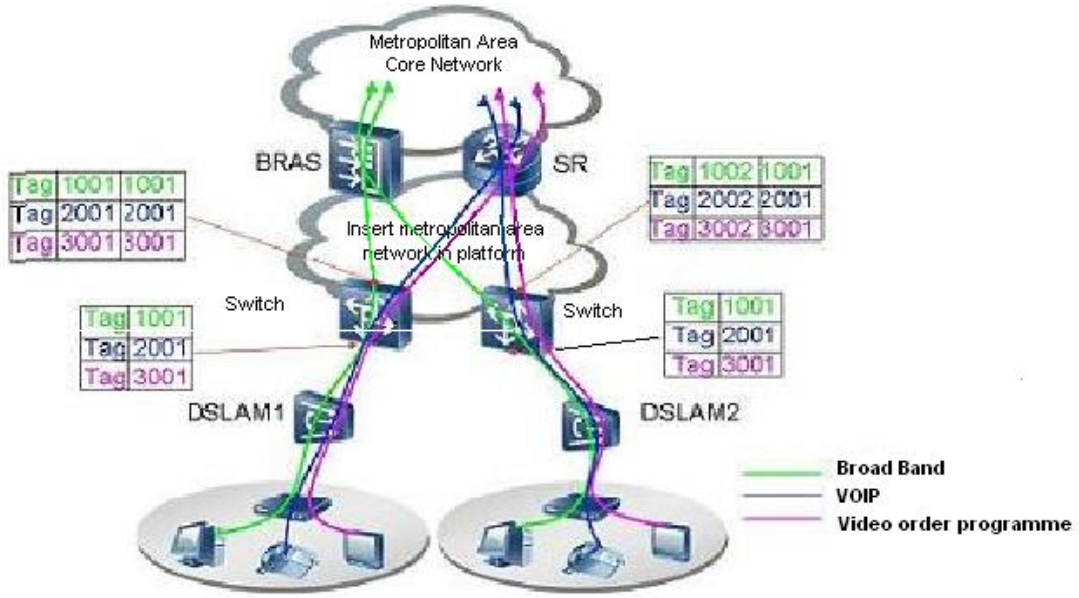


Figure 6-263 Flexible QinQ application topology

As shown in the figure, the first user is assigned three VLANs that the tag values are 1001, 2001, 3001 respectively in DSLAM1. VLAN1001 corresponds to Broad Band Network, VLAN2001 corresponds to VOIP, VLAN3001 corresponds to VOD. After the downlink port enables flexible QinQ function, the packets will be packed with different external tags according to VLAN ID of users. The packet with tag 1001 will be packed an external tag 1001 directly(This tag is unique in public network), enter Broad Band Network-VLAN1001 and classfied to BRAS device. The packet with tag 2001(or 3001) will be packed an external tag 2001(or 3001) and classfied to SR device according to the flow rules. The second user can be assigned different VLAN tags for different VLANs in DSLAM2. Notice: The assigned VLAN tag of the second user may be same with the first user and the packet with tag will be also packed an external tag. In the above figure, the external tag of the second user is different to the first user for distinguishing DSLAM location and locating the user finally.

The configuration in the following:

If the data flow of DSLAM1 enters the switch's downlink port1, the configuration is as follows:

```
Switch(config)#class-map c1
Switch(config-classmap-c1)#match vlan 1001
Switch(config-classmap-c1)#exit
Switch(config)#class-map c2
Switch(config-classmap-c2)#match vlan 2001
Switch(config-classmap-c2)#exit
Switch(config)#class-map c3
Switch(config-classmap-c3)#match vlan 3001
Switch(config-classmap-c3)#exit
Switch(config)#policy-map p1
Switch(config-policymap-p1)#class c1
Switch(config-policymap-p1-class-c1)#    set s-vid 1001
Switch(config-policymap-p1)#class c2
Switch(config-policymap-p1-class-c2)#    set s-vid 2001
```

```
Switch(config-policymap-p1)#class c3
Switch(config-policymap-p1-class-c3)#    set s-vid 3001
Switch(config-policymap-p1-class-c3)#exit
Switch(config-policymap-p1)#exit
Switch(config)#interface ethernet 1/0/1
Switch(config-if-ethernet1/0/1)#service-policy p1 in
```

If the data flow of DSLAM2 enters the switch's downlink port1, the configuration is as follows:

```
Switch(config)#class-map c1
Switch(config-classmap-c1)#match vlan 1001
Switch(config-classmap-c1)#exit
Switch(config)#class-map c2
Switch(config-classmap-c2)#match vlan 2001
Switch(config-classmap-c2)#exit
Switch(config)#class-map c3
Switch(config-classmap-c3)#match vlan 3001
Switch(config-classmap-c3)#exit
Switch(config)#policy-map p1
Switch(config-policymap-p1)#class c1
Switch(config-policymap-p1-class-c1)#    set s-vid 1002
Switch(config-policymap-p1)#class c2
Switch(config-policymap-p1-class-c2)#    set s-vid 2002
Switch(config-policymap-p1)#class c3
Switch(config-policymap-p1-class-c3)#    set s-vid 3002
Switch(config-policymap-p1-class-c3)#exit
Switch(config-policymap-p1)#exit
Switch(config)#interface ethernet 1/0/1
Switch(config-if-ethernet1/0/1)# service-policy p1 in
```

6.15.4 Flexible QinQ Troubleshooting

If flexible QinQ policy can not be bound to the port, please check whether the problem is caused by the following reasons:

- ☞ Make sure flexible QinQ whether supports the configured class-map and policy-map
- ☞ Make sure ACL includes permit rule if the class-map matches ACL rule
- ☞ Make sure the switch exists enough TCAM resource to send the binding
- ☞ Priority of flexible QinQ and vlan ingress filtering for processing packets is: flexible QinQ > vlan ingress filtering

6.16 MPLS QoS

6.16.1 MPLS QoS Introduction

The exp segment of MPLS（MultiProtocol Label Switch）provides the support for QoS, and hence a better service for the network communication.

6.16.1.1 MPLS QoS Terms

CoS：Class of Service, the class information carried in L2 802.1Q frames. It takes up 3 bits in the Tag segment of the frame header, and is called the user priority, ranging from 0 to 7.



Figure 6-274 The CoS Priority

DSCP：Differentiated Services Code Point, the class information carried in L3 IP headers. It takes up 6 bits, ranging from 0 to 63, and is downward compatible with IP Precedence.



Figure 6-285 The MPLS EXP Priority

A segment in MPLS messages presenting the service class of MPLS messages. It takes up 3 bits, ranging from 0 to 7.

Internal DSCP: the internal priority configuration of the switch, used to distinguish the priorities of the switch internal data messages, ranging from 0 to 63.

In-Profile: we call the flow within the range specified by the QoS monitor policy (the bandwidth or burst value) In-Profile.

Out-of-Profile: we call the flow exceeding the range specified by the QoS monitor policy (the bandwidth or burst value) Out-of-Profile.

6.16.1.2 The Realization of MPLS QoS

To realize QoS of L3 switch software, a universal and mature reference model is a prerequisite. QoS can’t create any new bandwidth, but it can adjust and configure the existing bandwidth resource to achieve the maximum efficiency. A complete applicable QoS can fully control and manage the network data transmission.

The MPLS QoS based on differentiated services will specify a priority for every packet at the entrance of the network. Such class information will be stored in the exp filed of the label. MPLS QoS provides same services to packets at the same priority level, and different services for packets with different priority. The switches or routers supporting MPLS QoS can provide different bandwidth to packets according to their class information, overwrite the class information of packets according to the monitor policy configuration and even drop some low-level packets when the bandwidth resource is tight.

6.16.2 MPLS QoS Configuration

The configuration task sequence of MPLS QoS is as follows:

1. Configure the class map

   After creating a class rule, such as matching according to exp, the switch will treat data flow of different classes with different policies.

2. Apply MPLS QoS to the port

   Set the trust mode of the interface as exp, or bind the policy. The polity can only take effect on a specific interface after being bound to the latter.

3. Configure the mapping relationship of MPLS QoS

   Configure the mapping from exp to internal priority, and the mapping from drop precedence and internal priority to exp.

4. Display the mapping relationship of MPLS QoS

1. Configure the match rule of the class map as exp

Command	Explanation
Global Configuration Mode	
match exp <exp-list> **no match exp**	Configure the match standard in class map, the no command deletes the specific match standard.

2. Configure trust exp

Command	Explanation
Port Configuration Mode	
mls qos trust exp **no mls qos trust**	Set the switch port to trust exp; the no operation will disable this trust state of the switch port. This command is not supported by switch.

3. Configure the MPLS QoS mapping

Command	Explanation
Global Configuration Mode	
mls qos map {exp-intp <intp1..intp8> / exp-dp *<dp1..dp8>*} **no mls qos map {exp-intp | exp-dp}** **mls qos map intp-exp <exp1..exp8>** **no mls qos map intp-exp**	Set the mapping from exp to internal priority, exp to drop priority and internal priority to exp.

4. Display the mapping relationship of MPLS QoS

Command	Explanation
Admin Mode	
show mls qos maps [exp-intp | exp-dp | intp-exp	Display the mapping relationship of MPLS

|]	QoS.

6.16.3 MPLS QoS Examples

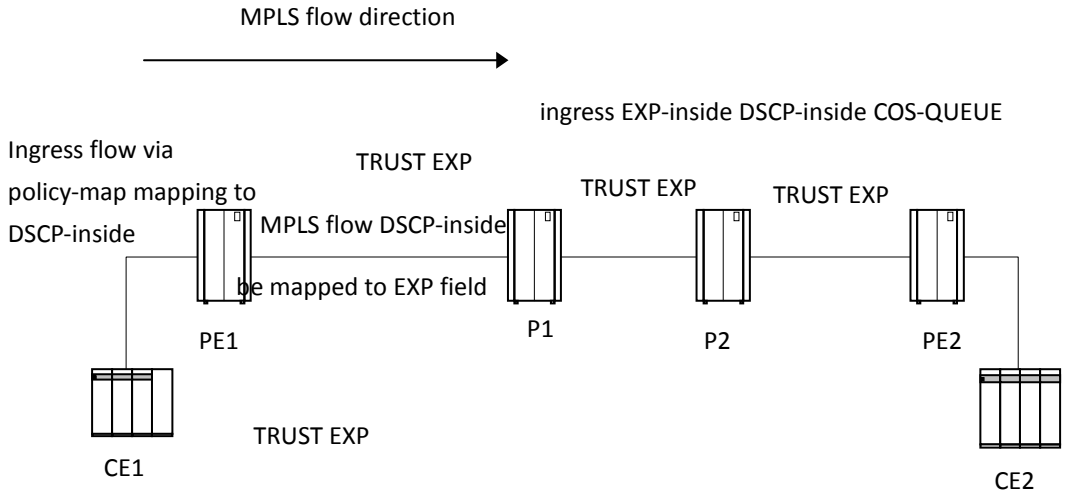


Figure 6-296 MPLS QoS Example

According to the diff-serv QOS model, the edge switch will classify the flow, and the core switch will forward the data packets according to their classes. As demonstrated in the above figure, the edge switch PE classifies the data flow according to the policy map, and store the result class in MPLS messages. The following switches P and PE, which are in the state of “trust EXP”, will forward the flow.

Configuration Examples:

Assume that normal data flows enter PE1 via vlan10, voip flows enter PE1 via vlan100, and the flows enter through Ethernet 1/0/1 and leave from Ethernet 1/0/2.

PE1：

```
Switch#config
Switch(config)#class-map voip
Switch(Config-ClassMap-voip)#match vlan 100
Switch(Config-ClassMap-voip)# exit
Switch(config)#class-map data
Switch(Config-ClassMap-data)#match vlan 10
Switch(Config-ClassMap-data)# exit
Switch(config)#policy-map p1
Switch(Config-Policy Map-p1)#class voip
Switch(Config-Policy Map-p1-Class-c1)#set Internal-Priority 1
Switch(Config-Policy Map-p1-Class-c1)#exit
Switch(Config-Policy Map-p1)#class data
Switch(Config-Policy Map-p1-Class-c1)#set Internal-Priority 0
Switch(Config-Policy Map-p1-Class-c1)#exit
Switch(Config-Policy Map-p1)#exit
```

Switch(config)#interface ethernet 1/0/1
Switch(Config-If-Ethernet1/0/1)#service-policy input p1
Switch(config)#interface ethernet 1/0/2
Switch(Config-If-Ethernet1/0/2)# mls qos queue wrr weight 1 2 3 4 5 6 7 8
Data flows, whose internal priority is 0, leave from queue 0 according to the default Int-Prio-TO-QUEUE, and EXP, DSCP and COS fields of the data packets are set as 0 according to the default Int-Prio-TO-(EXP, DSCP, COS) mapping.
Voip flows, whose internal priority is 1, leave from queue 1 according to the default Int-Prio-TO-QUEUE, and EXP, DSCP and COS fields of the data packets are set as 1 according to the default Int-Prio-TO-(EXP, DSCP, COS) mapping.
P1, P2, PE2: will be forwarded according to their classes, all flows will enter through Ethernet 1/0/1 and leave from Ethernet 1/0/2.
Switch#config
Switch(config)#interface ethernet 1/0/1
Switch Config-If-Ethernet1/0/1)#
Switch(config)#interface ethernet 1/0/2
Switch(Config-If-Ethernet1/0/2)#mls qos queue wrr weight 1 2 3 4 5 6 7 8
Data flows, whose EXP is 0, egress from queue 0 according to the default EXP-TO-(Int-Prio, Drop-Prec) and Int-Prio-TO-QUEUE, and EXP, DSCP and COS fields of the data packets are set as 0 according to the default Int-Prio-TO-(EXP, DSCP, COS) mapping.
Voip flows, whose EXP is 1, egress from queue 0 according to the default EXP-TO-(Int-Prio, Drop-Prec) and Int-Prio-TO-QUEUE, and EXP, DSCP and COS fields of the data packets are set as 1 according to the default Int-Prio-TO-(EXP, DSCP, COS) mapping.

6.16.4 MPLS QoS Troubleshooting Help

- The MPLS should be enabled on the switch port otherwise the MPLS QoS will be unavailable.
- After passing an interface with MPLS QoS enabled, the cos value of MPLS messages will be set to 0 while dscp will stay the same.

6.17 Egress Queue Scheduling

6.17.1 Introduction to Egress Queue Scheduling

When forwarding packets results in congestion, with egress queue scheduling, high priority queue is served prior to low priority queue according to packet's priority, so as to implement QoS aim. The chip supports the scheduling arithmetics of SP, RR, WRR, and WDRR.Etc. According to different scheduling levels, there is single-stage scheduling and multi-stage scheduling. Under the single-stage scheduling, the data packet will come out from the egress after it wined in the

queues competition. Under the multi-stage scheduling, the data packet will enter into the last level of scheduling node to join the competition after it wined in the queues competition, the physical port is the highest level of the scheduling node and only the data packet which wined in the highest level of scheduling node will be forwarded out from the port. This device supports the three-level scheduling.

6.17.1.1 Egress Queue Scheduling Terms

Scheduling: QoS egress action. Add the packets to the corresponding egress queue according to the internal priority. And then decide sending and dropping according to Drop Precedence, sending algorithm and queue weight of egress queue.
Internal Priority: The internal priority setting of the switch chip, its valid range relates with the chip, its shortening is Int-Prio or IntP.
L2 CoS: Class of Service, the classification information carried by Layer 2 802.1Q frames, taking 3 bits of the Tag field in frame header, is called user priority in the range of 0 to 7.
ETS: Enhanced Transmission Selection, it processes multi-level scheduling for different traffic. (IEEE 802.1Qaz)
UC: known unicast packet, it is the only one of the forwarding destination port.
MC: non-unicast packet, it mainly includes the unknown unicast, multicast, broadcast, and mirror image.etc.

6.17.1.2 Egress Queue Scheduling Implement

Egress queue scheduling mechanism supports three-level scheduling, besides, UC packets and MC packets can configure scheduling mode respectively, WRR and WDRR scheduling modes can configured scheduling weight respectively. In queue scheduling, UC packets include unicast packets, MC packets include broadcast packets, multicast packets (Except the Multicast traffic which is expected to have a single destination in the upstream direction), monitor data packets and DLF packets, and so on.

For packet flow, packets should distinguish UC packets or MC packets at first, and set the mapping from CoS to queue in COS_MAP table, form the corresponding mapping queue (at present, unicast packets with cos values from 0 to 7 are mapped to UC queues with number from 1 to 8, non-unicast packets with cos value of 0 are mapped to queue1 of MC, non-unicast packets with cos values from 1 to 3 are mapped to queue2 of MC, non-unicast packets with cos values from 4 to 6 are mapped to queue3 of MC, non-unicast packets with cos value of 7 are mapped to queue4 of MC) to process three-level scheduling. Scheduling flow is shown in the following:

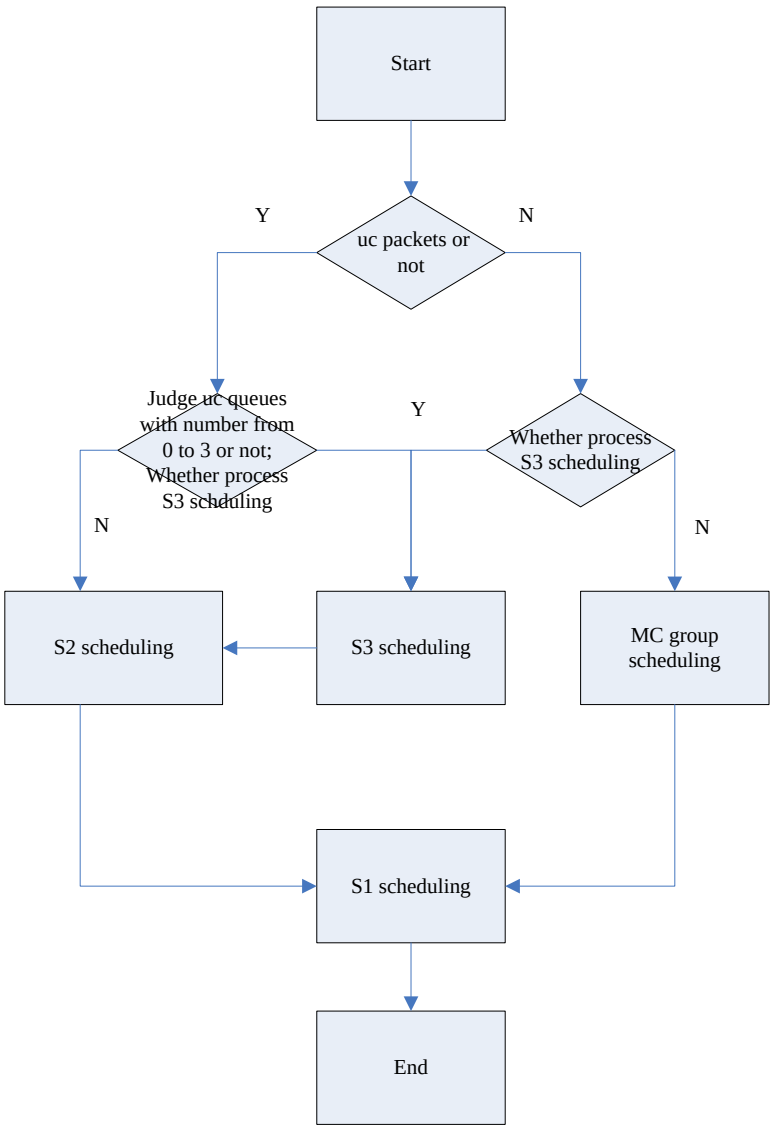


Figure 6-47 Three-level scheduling flow chart

In the above figure, packets will pass S3 scheduling, S2 scheduling (MC Group) and S1 scheduling in turn, S3 is able to schedule four queues with MC packets and 0-3 queues with number from 0 to 3 about UC packets, set mapping from these queues to S2 scheduling node. S2 scheduling node includes three scheduling nodes, their scheduling results and other queues (UC queues, QM and SC queues of MC) enter S1 scheduling node. Besides, S3 is set as RR scheduling mode, S2 (including MC_GROUP) is set as WRR scheduling mode, S1 is set as SP scheduling mode by default. To be mentioned, only four MC queues can be scheduled to S3 scheduling, but all MC queues are scheduled to MAC-GROUP to directly enter S1 scheduling.

6.17.1.3 Basic Egress Queue Scheduling Model

There are two scheduling modes supported by port. One is the normal scheduling mode which is compatible with the previous command mode (it is called as the normal mode), the other is ETS scheduling mode.

Switch uses the normal mode as the default scheduling mode, its mechanism is shown in the

following, thereinto S3 scheduling node uses RR scheduling algorithm, S2.1 uses WRR scheduling algorithm.

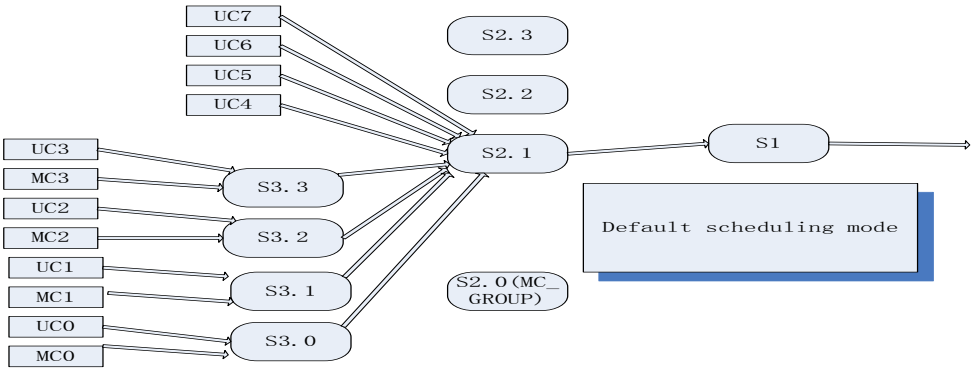


Figure 6-48 Default scheduling mode

To provide many selections, ETS scheduling mode includes UC-MC mode and advanced ETS mode. UC-MC scheduling mechanism is shown in the following, thereinto S2.1 and mc-group nodes use WRR scheduling algorithm, S1 node uses SP scheduling algorithm. Under the UC-MC mode, S1 node adopts SP scheduling arithmetic always, so there is the strict priority relation. User can select to schedule UC or MC flow first.

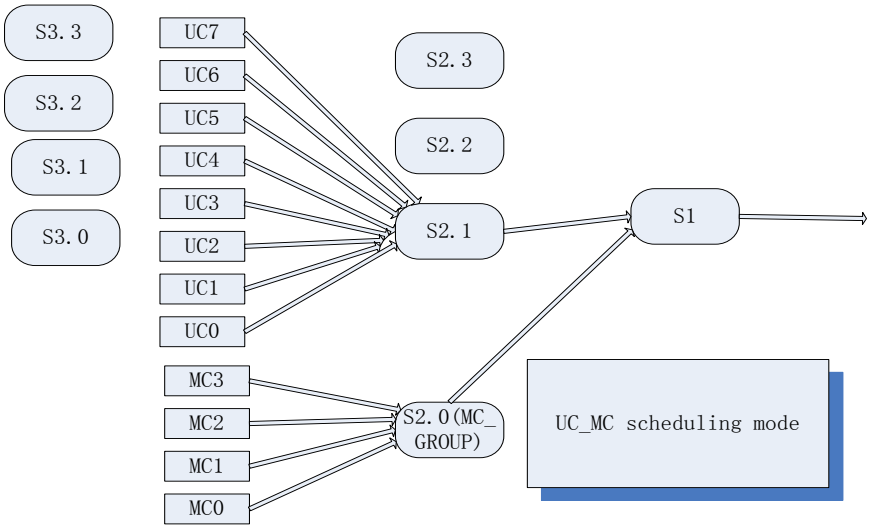


Figure 6-49 UC-MC scheduling mode

Advanced ETS scheduling mechanism is shown in the following, a UC queue, a MC queue or a scheduling node can be mapped to one queue of upper node only, at the same time, there is only one input source for each queue of any scheduling node. UC7~UC4 can be mapped to S2.1/S2.2/S2.3 node only, UC3~UC0 can be mapped to S2.1/S2.2/S2.3 node or S3.3~S3.0 node (If they are mapped to S3 scheduling node, UC3~UC0 must be mapped S3.3~S3.0 respectively). MC3~MC0 can be mapped to S3 scheduling node or S2.0 (MC_GROUP) node (If they are mapped to S3 scheduling node, MC3~MC0 must be mapped to queue with number 0 on S3.3~S3.0 respectively). S3.3~S3.0 only be mapped to S2.1/S2.2/S2.3, and S2.3~S2.0 only be mapped to queues with number from 0 to 3 in S1 scheduling node. Under the advanced ETS mode, the chip provides the most flexibility, user can configure the levels of flow joining scheduling in the range that chip can bear.

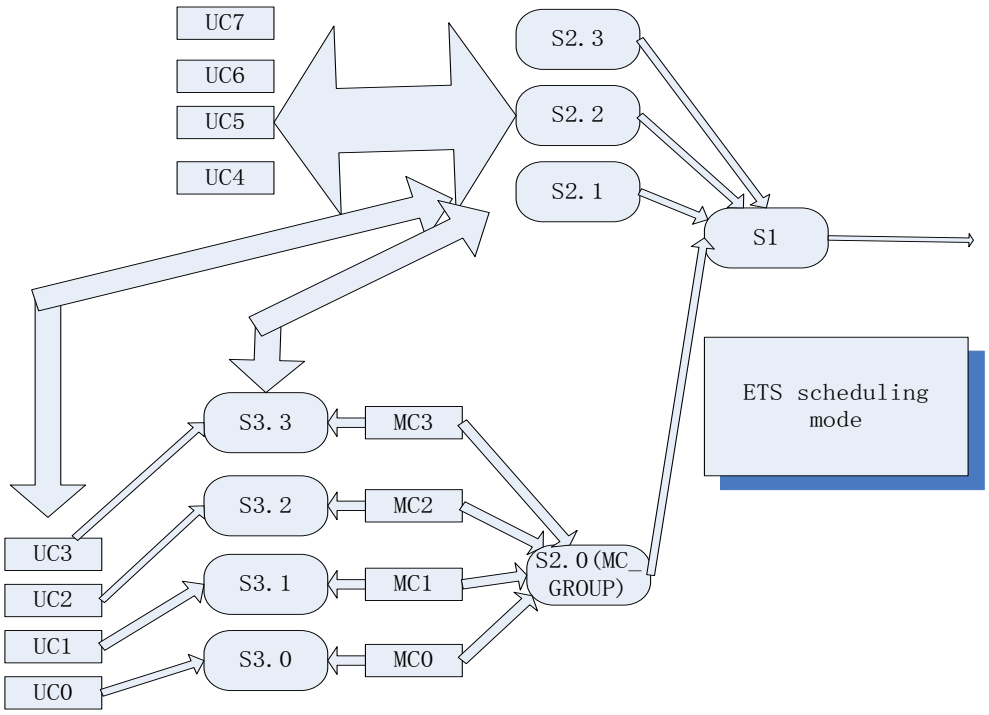


Figure 6-300 Advanced ETS scheduling mode

6.17.2 Egress Queue Scheduling Configuration

Egress queue scheduling configuration task list:

1. Configure scheduling mode for ports

   Configure egress queue scheduling as normal mode, UC-MC mode or advanced ETS mode.

2. The scheduling policy table can be configured if enabled advanced ETS mode on the port

   Create ETS scheduling policy table under the global mode. This table is configured globally, it can be applied onto multi-ports.

3. Configure scheduling node

   Create the scheduling node under the scheduling policy table mode. Currently, S1 node only can configure one node, named S1.1. 4 can be configured at most in S2, named S2.1~S2.4, S2.1 only can be used in scheduling MC flow, S2.2~S2.4 cannot be used in scheduling MC flow. 4 can be configured at most in S3, named S3.1~S3.4.

4. Configure the data source of every subqueue of the scheduling node

   User can define the the data source of every subqueue of the scheduling node. The multi-stage scheduling can be reflected through this command.

5. Configure the maximum and minimum bandwidth of scheduling arithmetic, scheduling weight, scheduling node

   Under the scheduling node mode, user can configure the scheduling arithmetic used in scheduling node and configure the queue weight of every subqueue. User can also configure the integral maximum and minimum bandwidth of this scheduling node. Actually, there are no commands that can configure the maximum and minimum bandwidth of every subqueue. So configuring the integral maximum and minimum bandwidth of the scheduling node is equivalent

to configuring the maximum and minimum bandwidth of the last level node of this scheduling node.

6. Configure the maximum and minimum bandwidth of multicast and unicast queues

Configure the maximum and minimum bandwidth of multicast and unicast queues.

7. Bind scheduling policy map to port's egress

A policy will only take effect on a port when it is bound to that port. The policy only can be bond to the port which has enabled ETS but not enabled UC-MC mode.

Configure queue scheduling when ETS is not enabled on a port

When port does not enable ETS, the corresponding queue scheduling configuration is also configured in normal node, such as queue scheduling algorithm, queue weight.

Configure bandwidth limit of queue

When port enables ETS, MC and UC queue bandwidth limit can be configured. When port does not enable ETS, UC queue bandwidth limit can be configured.

Show queue scheduling command

Check queue scheduling configuration of port by show command.

1. Configure scheduling mode for ports

Command	Explanation
Port mode	
mls qos ets enable **no mls qos ets enable**	Enable ETS queue scheduling mode for a port, the no command disables ETS mode.
mls qos queue {uc | mc} higher **no mls qos queue {uc | mc} higher**	Configure the priority of UC or MC, the no command disables the priority configuration.

2. Configure ETS scheduling policy map

Command	Explanation
Global mode	
mls qos schedule policy <polcicyName> **no mls qos schedule policy <polcicyName>**	Create a scheduling policy and enter the scheduling policy mode, the no command deletes the specified scheduling policy.
Scheduling policy mode	
mls qos schedule level <levelID> node <nodeID> **no mls qos schedule level <levelID> node<nodeID>**	Create a scheduling node and enter the corresponding scheduling node mode, the no command cancels the specified scheduling node configuration.
Scheduling node mode	
mls qos schedule queue <queueID> input {{UC <ucID>} | {MC <mcID>} | {node <nodeID>}} **no mls qos schedule queue <queueID> input {{UC <ucID>} | {MC <mcID>} | {node <nodeID>}}**	Configure the input data source of a queue on a scheduling node, the no command cancels the operation.

mls qos bandwidth <min-bandwidth> <max-bandwidth> **no mls qos bandwidth <min-bandwidth> <max-bandwidth>**	Configure the bandwidth limit for a scheduling node, the no command cancels the configuration.
mls qos schedule algorithm {sp | wdrr | wrr} **no mls qos schedule algorithm**	Configure queue scheduling algorithm for a scheduling node, the no command restores the default scheduling algorithm.
mls qos schedule {wrr | wdrr} weight <weight0..weight17> **no mls qos schedule {wrr | wdrr} weight**	Configure queue scheduling weight when the port does not enable ETS, the no command restores the default weight.

3. Apply egress scheduling policy to port

Command	Explanation
Port mode	
mls qos schedule policy bind <policyName> **no mls qos schedule policy bind <policyName>**	Bind a scheduling policy to a port, the no command cancels the binding between the scheduling policy and the port.

4. Configure queue scheduling when ETS is not enabled on a port

Command	Explanation
Port mode	
mls qos queue algorithm {sp | wrr | wdrr} **no mls qos queue algorithm**	Configure queue scheduling algorithm for a port in normal scheduling mode, the no command restores the default scheduling algorithm.
mls qos queue {wrr | wdrr} weight <weight0..weight17> **no mls qos queue {wrr | wdrr} weight**	Configure queue scheduling weight when the port does not enable ETS, the no command restores the default weight.

5. Configure bandwidth limit of queue

Command	Explanation
Port mode	
mls qos mc queue <queue-id> bandwidth <min-bandwidth> <max-bandwidth> **no mls qos mc queue <queue-id> bandwidth**	Configure the bandwidth limit for a MC queue after a port enables ETS. The no command cancels the configuration.
mls qos queue <queue-id> bandwidth <minimum-bandwidth> <maximum-bandwidth> **no mls qos queue <queue-id> bandwidth**	Configure the bandwidth limit for UC queue on a port. The no command cancels the configuration.

6. Show queue scheduling command

Command	Explanation

Admin mode, global mode, port mode	
show mls qos ets interface [<interface-name> | <interface-list>]	Show whether the interface enables ETS scheduling.
show mls qos schedule interface [<interface_name> | <interface_list>]	Show queue scheduling configuration for the interfaces.

6.17.3 Egress Queue Scheduling Examples

Example:

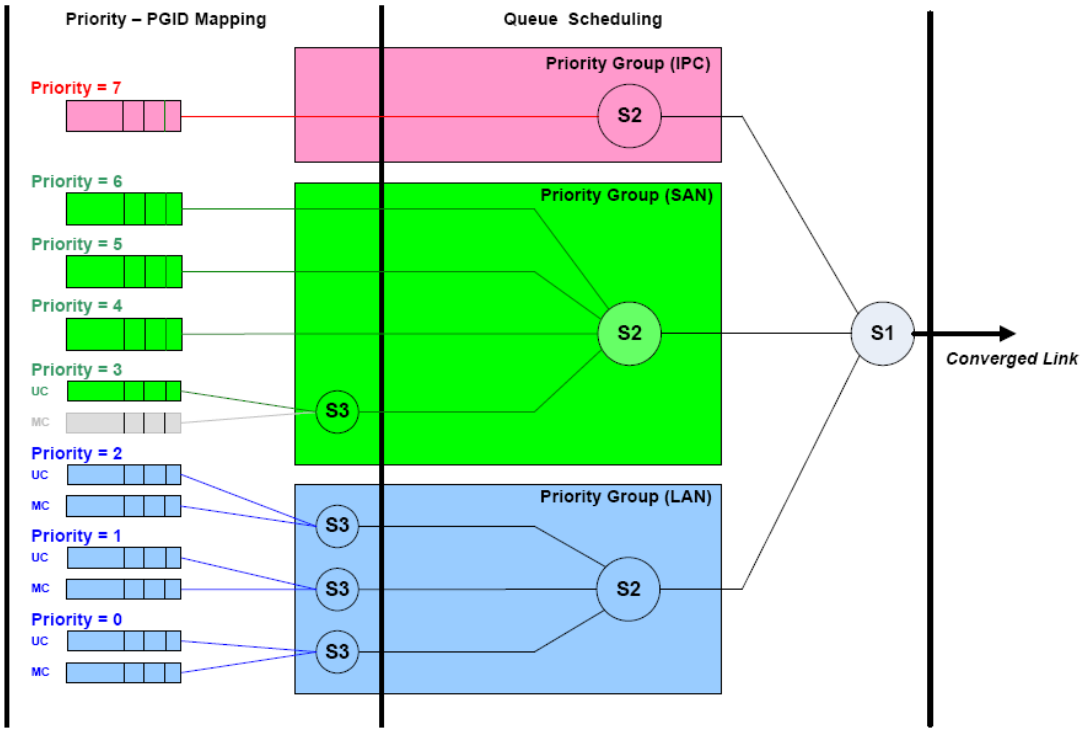


Figure 6-51

To distinguish the priority of different data streams and schedule them in DCB (DATA CENTER BRIDGE) environment, there are three kinds of data stream, they are LAN, SAN (Storage Area Network), IPC (Inter Process Communication). Thereinto, IPC priority is the highest, the second is SAN data (usually, it is UC data). Configuration is shown in figure, UC7 packets are the high priority IPC packets with the low latency, such as voice, they enter S2.2 scheduling directly. UC4-6 is SAN packets, UC3 and MC3 belong to LAN packets, they enter S3.3 scheduling node directly to be scheduled to S2.1 scheduling. UC0-2 and MC0-2 will enter S3.0~S3.2 scheduling node, so as to be scheduled by S2.1 scheduling node. Besides, S1 scheduling is configured as SP+WDRR scheduling mode, namely, IPC packets can be scheduled by queue scheduling firstly as long as they exist, other packets are scheduled with WDRR scheduling algorithm according to the configured weight.

Configuration procedure:

Switch(config)#mls qos schedule policy p

Switch (config-sche-policy-p)#mls qos schedule level 3 node 1

Switch (config-sche-policy-p-lel3nod1)#mls qos schedule queue 1 input mc 1

```
Switch (config-sche-policy-p-lel3nod1)#mls qos schedule queue 2 input uc 1
Switch (config-sche-policy-p-lel3nod1)#exit
Switch (config-sche-policy-p)#mls qos schedule level 3 node 2
Switch (config-sche-policy-p-lel3nod2)#mls qos schedule queue 1 input mc 2
Switch (config-sche-policy-p-lel3nod2)#mls qos schedule queue 2 input uc 2
Switch (config-sche-policy-p-lel3nod2)#exit
Switch (config-sche-policy-p)#mls qos schedule level 3 node 3
Switch (config-sche-policy-p-lel3nod3)#mls qos schedule queue 1 input mc 3
Switch (config-sche-policy-p-lel3nod3)#mls qos schedule queue 2 input uc 3
Switch (config-sche-policy-p-lel3nod3)#exit
Switch (config-sche-policy-p)#mls qos schedule level 3 node 4
Switch (config-sche-policy-p-lel3nod4)#mls qos schedule queue 1 input mc 4
Switch (config-sche-policy-p-lel3nod4)#mls qos schedule queue 2 input uc 4
Switch (config-sche-policy-p-lel3nod4)#exit
Switch (config-sche-policy-p)#mls qos schedule level 2 node 2
Switch (config-sche-policy-p-lel2nod2)#mls qos schedule queue 1 input node 1
Switch (config-sche-policy-p-lel2nod2)#mls qos schedule queue 2 input node 2
Switch (config-sche-policy-p-lel2nod2)#mls qos schedule queue 3 input node 3
Switch (config-sche-policy-p-lel2nod2)#exit
Switch (config-sche-policy-p)#mls qos schedule level 2 node 3
Switch (config-sche-policy-p-lel2nod3)#mls qos schedule queue 1 input node 4
Switch (config-sche-policy-p-lel2nod3)#mls qos schedule queue 2 input uc 5
Switch (config-sche-policy-p-lel2nod3)#mls qos schedule queue 3 input uc 6
Switch (config-sche-policy-p-lel2nod3)#mls qos schedule queue 4 input uc 7
Switch (config-sche-policy-p-lel2nod3)#exit
Switch (config-sche-policy-p)#mls qos schedule level 2 node 4
Switch (config-sche-policy-p-lel2nod4)#mls qos schedule queue 1 input uc 8
Switch (config-sche-policy-p-lel2nod4)#exit
Switch (config-sche-policy-p)#mls qos schedule level 1 node 1
Switch (config-sche-policy-p-lel1nod1)#mls qos schedule algorithm wdrr
Switch (config-sche-policy-p-lel1nod1)#mls qos schedule wdrr weight 1 2 4 0 0 0 0 0
Switch (config)#interface ethernet 1/0/2
Switch (config-if-ethernet1/0/2)#mls qos ets enable
Switch (config-if-ethernet1/0/2)#mls qos schedule policy bind p
```

6.17.4 Egress Queue Scheduling Troubleshooting

- Whether enable ETS of the port, the egress flow will always be scheduled according to UC and MC.
- When configuring UC/MC preference of port, it must firstly enable ETS for port.
- MC3~MC0 can be mapped to S3 scheduling node or S2.0 (MC_GROUP) scheduling node. If they are mapped to S3 scheduling node, MC3~MC0 must be mapped to queue with number 0 of S3.3~S3.0 respectively.

- If ETS is not enabled for port, mls qos queue <queue-id> bandwidth <minimum-bandwidth> <maximum-bandwidth> takes effect to UC queues. To be mentioned, only bandwidth limit of UC5-8 queues take effect due to chip disfigurement in this condition.
- Egress queue scheduling supports the dynamic configuration. In global mode, the relative operation of multi-level queue scheduling will prompt the information (existing the unsupported line card) as long as the switch exists the line card without multi-level queue scheduling.
- We suggest user to use **show mls qos schedule interface** command to check the scheduling configuration message of some ports. Even the same scheduling policy is applied on multi-ports, the actual port scheduling configuration message may be different because of the different functions of the port.

6.18 Vlan-shaping

6.18.1 Introduction to Vlan-shaping

VLAN Shaping is the differentiated service to achieve flow plastic for the specified VLAN. In Vlan Shaping, vlan id in the packet is equivalent to the number of the differentiated service, it is the core difference compared with the traditional flow plastic. The traditional flow plastic is based on the port and the queue, and the foundation to achieve differentiated service is the field of cos/dscp/exp of the packet mainly. So the traditional QOS is based on the business mainly. It cannot provide independent QOS service for different users under the current application of multi-user and multi-service sharing data link.

6.18.1.1 Vlan-shaping Terms

QoS：Quality of Service provides a guarantee for service quality of consistent and predictable data transfer service to fulfill program requirements. QoS cannot generate new bandwidth but provides more effective bandwidth management according to the application requirement and network management.

VID： VLAN ID is in Vlan Tag and used for distinguishing the 12bit of vlan, the range is 1 to 4094.

Queue： It is the egress scheduling queue of Cos.

EUC Queue: Extended Unicast Queue, is the extend unicast queue and used for Vlan flow scheduling queue. So in this text, euc queue is the same as vlan shaping queue.

Forwarding-class: Forwarding class is used to sign a class of flow and it is the original input source in multi-level scheduling. If the multi-level scheduling is treated as a tree, S1 is equivalent to the root, S2 and S3 are equivalent to branch node, the original input queue (unicast queue, multicast queue, extend unicast queue) is equivalent to the leaf node and it is the forwarding class. Currently, the forwarding class is specific to the extend unicast queue (vlan shaping).

Forwarding-class_match: It is the matching rules of forwarding class and used to classify the flow. The flow in accordance with the classification rules will enter the original queue that the forwarding class corresponded to participate the egress queue scheduling. Currently, the

forwarding class is only matching the vlan id in outer tag.

Forwarding-profile： It is used to configure the action related to the queue scheduling for some flow configuration, such as queue scheduling algorithm, queue scheduling weight, the minimum and maximum bandwidth and so on. Currently, the forwarding class can only configure the minimum and maximum bandwidth and quote some drop policy. A forwarding policy can be quoted by a forwarding class, it means scheduling the action for the queue configured by flow of the forwarding class.

Drop-profile: It is used to configure the action related to the queue drop for some flow configuration, such as configuring the parameter of wred. A drop policy can be quoted by a forwarding policy, it means dropping queue according to the wred algorithm configured by the drop policy.

6.18.1.2 Vlan-shaping Implementation

The following is the schematic of vlan flow egress scheduling plastic:

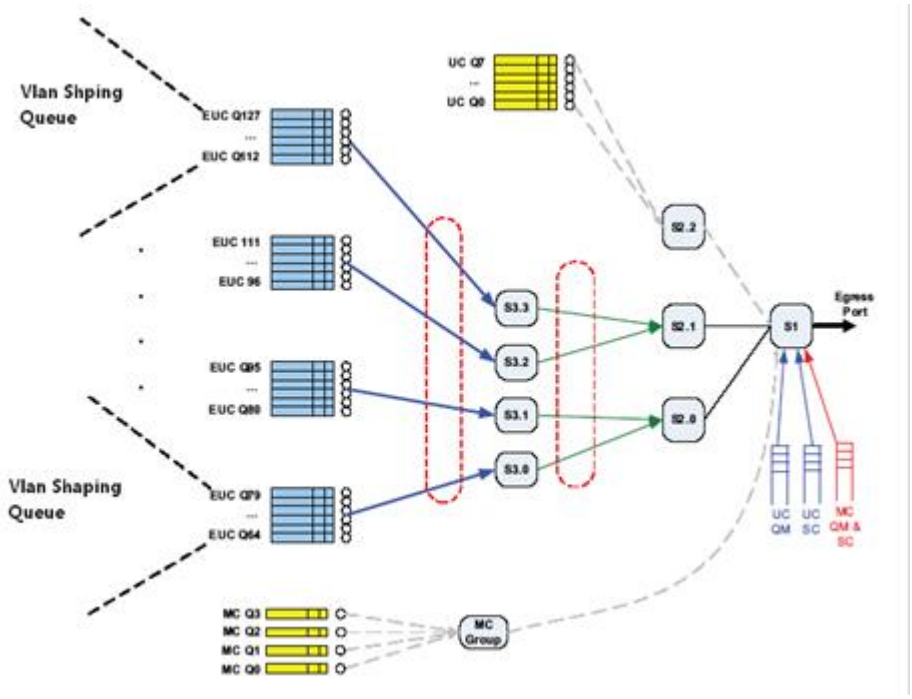


Figure 6-312 vlan-shaping scheduling achieving

In the figure above, the switch supports 64 vlan shaping. The 64 queues are into 4 groups and there are 16 queues per group. There are 4 nodes of S3 and in every node there is a queue group of vlan shaping. Every node of S3 supports 18 queues, the first two are MC and UC and the last 16 are the queues of vlan shaping. So vlan shaping is the same as a mechanism of S3 essentially.

In the occasion of enabling the function of vlan shaping, the packet is put into the specified queue of vlan shaping to participate scheduling according to vlan id in tag. The cos/dscp/exp of the original packet cannot decide the number of the queue now. In the occasion of not enabling the function of vlan shaping, the packet decides the number of the queue according to the trust style of the port (trust cos/trust dscp/trust exp).

6.18.1.3 Basic QoS Model

VLAN Shaping model includes VLAN, CE, PE and PCN, the following is the basic concept:

VLAN: VLAN in VLAN Shaping means OUTER VLAN. The operators identify different users or business through Outer Vlan Id.

CE (Custom Edge): It is the users' side-edge equipment. The one side supports the uplink port from link to the network of operators and another side supports the downlink port from link to the network of users.

PE (Provider Edge): It is the operators' side-edge equipment. Its uplink port is connected to the network of operators and the downlink port supports the access to the network of users through connecting to the uplink port of CE.

PCN (Provider Core Network): It is the core network of the operators.

A typical application of vlan shaping is that support the specified QoS service for different vlan flow which comes from different users through vlan shaping on the uplink port of PE.

In the figure below, there are voice, data and video in the network. The requirement to the detention and the bandwidth of the network in these three kinds of business is different. The requirement of voice business to the bandwidth is not high, but it is very high to the detention of the network. The requirement to the detention and the bandwidth of the network of video business is high. The requirement to the detention and the bandwidth of the network of the common data is in between. The traditional QOS can support differentiation service for different businesses.

In practical application, the link access that operators support to the users is limited. Multiple users share one uplink access. In the figure bellow, there are two kinds of users and they located in VLAN2-199 and VLAN200-300. These users share one link access to go to the PE through the equipment convergence to CE.

For the operators, supporting different QOS services on the sharing link to different users can protect and utilize the hardware resource existing as much as possible. The operators have not to support different link accesses for different users. They support the relevant QOS with user on every link access. On this sharing link access, the only thing to do is to distinguish which user the flow comes from. Vlan id can sign this kind of users and support QOS on the uplink port of PE.

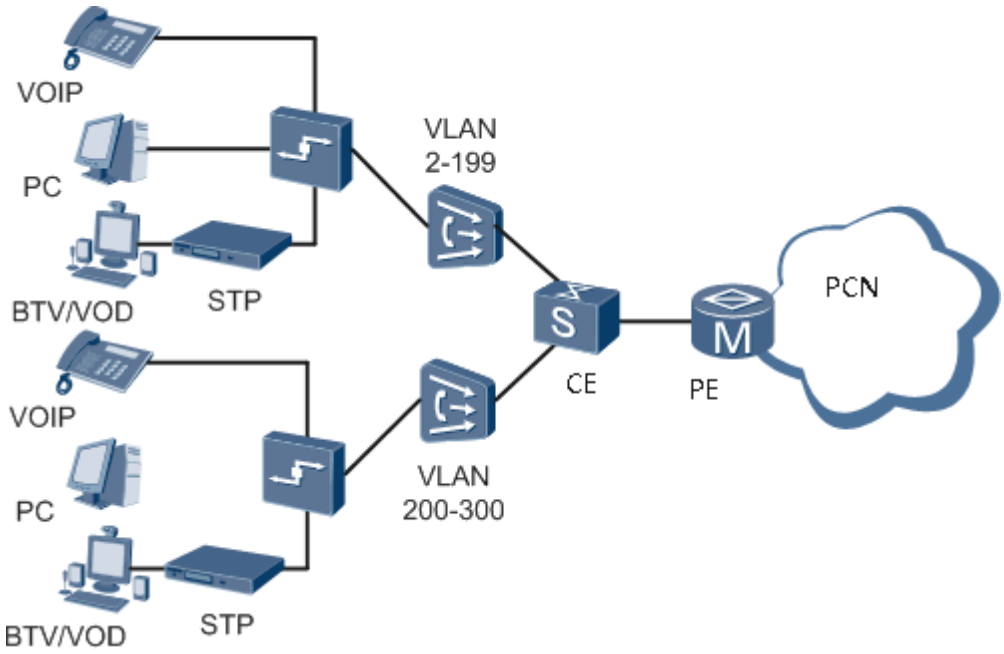


Figure 6-323 the typical application of vlan shaping under the application of multi-user and multi-service

6.18.2 Vlan-shaping Configuration Task List

The following is the vlan-shaping configuration task list:

Forwarding-class Configuration

Create a forwarding class and it can match different vlan. Then bind different forwarding class to different schedule policy and take different policies to the data flow in this forwarding class. This configuration is required.

Drop-profile Configuration

Create a drop policy to configure the parameters of WRED. This configuration is not required. Before enabling wred or using the parameter of wred, this configuration has not to be configured.

Forwarding- policy Configuration

Create a forwarding policy, it can configure the bandwidth and quote the drop policy. Then it can be binding to the schedule policy with the forwarding class to control the data flow in the forwarding class. This configuration is not required. If this configuration is not be configured, there is not the bandwidth restriction and the parameter of wred is default.

Schedule Policy Configuration

Create a schedule policy, and then map the forwarding class and forwarding policy to every queue of the four nodes of level3. Configure the queue scheduling algorithm and the queue weight. This configuration is required but quoting the forwarding policy is not required. If the forwarding policy is not quoted, there is not the bandwidth restriction and the parameter of wred is default.

Enable the Function of ets on the Port (Multi-level Scheduling)

Only enable the function of ets on the port, the custom schedule policy can be bond to the port. This configuration is required.

Utilize Vlan-shaping to the Port

Configure the binding policy of the port. Only the policy is bond to the specific port, the policy in this port can be effective. This configuration is required.

1. Forwarding-class Configuration

Command	Explanation
Global Mode	
mls qos forwarding-class <fc-name> [id <fc-id>] **no mls qos forwarding-class <fc-name>**	Create a forwarding class and enter the mode of forwarding class; the **no** command deletes the specified forwarding class.
Forwarding Class Mode	
match s-vid<svid-list> **no match s-vid[svid-list]**	Configure the matching standard of the forwarding class; the data flow can be classified according to vlan-id. The **no** command deletes the specified matching standard.

2. Forwarding-profile Configuration

Command	Explanation
Global Mode	
mls qos forwarding-profile <fp-name> [id <fp-id>] **no mls qos forwarding-profile <fp-name>**	Create a forwarding policy and enter the mode of forwarding policy. The **no** command deletes the specified forwarding policy.
Forwarding Policy Mode	
Bandwidth <minBandwidth> <maxBandwidth> **no bandwidth**	Configure the minimum and the maximum bandwidth in the forwarding policy. The **no** command deletes the bandwidth.
drop <dp-name> **no drop**	Quote the drop policy in the forwarding policy. The **no** command cancels the quotation.

3. Schedule Policy Configuration

Command	Explanation
Global Mode	
mls qos schedule policy <fp-name> **no mls qos schedule policy <fp-name>**	Create a schedule policy and enter the mode of schedule policy; The **no** command deletes the specified schedule policy.
Schedule Policy Mode	
mls qos schedule level <level-id> node <node-id> **no mls qos schedule level <level-id> node <node-id>**	Enter the configuration of schedule node; The **no** command deletes the specified node configuration.
Schedule Node Mode	
mls qos schedule algorithm {sp|wrr|wdrr}	Configure the queue scheduling

no mls qos schedule algorithm	algorithm. For the switch which supports this function, it should configure this command under the mode of S3 schedule node, the default is rr.
mls qos schedule {wrr|wdrr} weight <weight1…weight18> **no mls qos schedule {wrr|wdrr} weight**	Configure the queue weight of vlan shaping. For the switch which supports this function, it should configure this command under the mode of S3 schedule node; the node of S3.1 has 18 queues. The default of wrr is 1, 2, 3 and so on. The default of wdrr is 10, 20, 40, 80, 127, 127, 127 and so on.
mls qos bandwidth <minBandwidth> <maxBandwidth> **no mls qos bandwidth**	Configure the minimum and the maximum bandwidth in the forwarding policy. For the switch which supports this function, it should configure this command under the mode of S3 schedule node. The **no** command deletes the configuration of bandwidth restriction.
mls qos schedule queue <queueID> input forwarding-class <fc-name> [profile <fp-name>] **no mls qos schedule queue <queueID> input forwarding-class <fc-name>**	Configure the input source of the queue of schedule node as forwarding class. And it can assign the forwarding policy for forwarding class selectively. The **no** command deletes the binding configuration.

4. Drop Policy

Command	Explanation
Global Mode	
mls qos drop-profile <dp-name> [id <dp-id>] **no mls qos drop-profile <dp-name>**	Create the drop policy and enter the mode of drop policy. The no command deletes the specified drop policy.
Drop Policy Mode	
dp <dp> drop-startpoint <start> drop-endpoint <end> max-drop-rate <rate> **no dp [dp]**	Configure the parameter of wred in the drop policy including type of color, minimum value, maximum value and the maximum drop probability. The **no** command deletes the configuration of color and return to the default.

5. Show the Information of Vlan-shaping Configuration

Command	Explanation
Global Mode	
show mls qos vlan shaping capable ports	Show all the names of the ports which support vlan shaping now.
show mls qos forwarding-class [fc-name]	Show the information of the specified

	forwarding class.
show mls qos drop-profile [dp-name]	Show the information of the specified drop policy.
Show mls qos forwarding-profile [fp-name]	Show the information of the specified forwarding policy.
show mls qos schedule interface [<interface-name>|<interface-list>]	Show the information of vlan shaping on the specified port including the queue scheduling algorithm of vlan shaping, the weight of the queue and so on. The information of S3 shows the information of the queue of vlan shaping.

6.18.3 Vlan-shaping Examples

Case1:

Create the forwarding class of fc1 and fc2, configure the matching rules as vlan1, vlan2, vlan3 and vlan4 flow. Create the drop policy of dp1 and configure it. For the packet whose drop priority is 0, it begins to drop packets when the queue length is up to 20% of the maximum length. It begins to drop packets in full speed when the queue length is up to 40% of the maximum length, but the drop packets account for the proportion of all packets to be 80%. Create the forwarding policy of fp1, its minimum bandwidth pledge is 10kbps and its maximum bandwidth restriction is 200kbps. Associate the forwarding policy to the drop policy of dp1. Create the scheduling policy of p1 and configure its first queue's inputting source of node of S3.1 as fc1, associate it to the forwarding policy of fp1. Configure its second queue's inputting source of node of S3.1 as fc2. Configure the minimum bandwidth pledge is 64kbps and its maximum bandwidth restriction is 256kbps of node of S3.1 of vlan-shaping queue. Configure the wrr queue's weights of vlan-shaping queue of 1 to 8 as 1, 3, 5, 2, 7, 4, 9, 14. And configure the wrr queue's weights of vlan-shaping queue of 9 to 18 as 8, 3, 2, 14, 17, 4, 7, 0, 0, 0. At last, bind p1 to the port of ethernet1/0/1.

Case Configuration:

(config)#mls qos forwarding-class fc1
(config-forwarding-class-fc1)#match vlan 1-2
(config-forwarding-class-fc1)#exit
(config)#mls qos forwarding-class fc2
(config-forwarding-class-fc1)#match vlan 3-4
(config-forwarding-class-fc1)#exit
(config)# mls qos drop-profile dp1
(config-drop-profile-dp1)# dp 0 drop-startpoint 20 drop-endpoint 40 max-drop-rate 80
(config-drop-profile-dp1)#exit
(config)#mls qos forwarding-policy fp1
(config-forwarding-policy-fp1)# bandwidth 10 200
(config-forwarding-policy-fp1)#drop dp1

(config-forwarding-policy-fp1)#exit

(config)#mls qos schedule policy p1

(config-schedule-policy-p1)#mls qos schedule level 3 node 1

(config-schedule-policy-p1-lel3nod1)#mls qos schedule queue 3 input forwarding-class fc1 profile fp1

(config-schedule-policy-p1-lel3nod1)#mls qos schedule queue 4 input forwarding-class fc2

(config-schedule-policy-p1-lel3nod1)#mls qos bandwidth 64 256

(config-schedule-policy-p1-lel3nod1)#mls qos schedule algorithm wrr

(config-schedule-policy-p1-lel3nod1)#mls qos schedule wrr weight 1 3 5 2 7 4 9 14 8 3 2 14 17 4 7 0 0 0

(config-schedule-policy-p1-lel3nod1)#exit

(config-schedule-policy-p1)#exit

(config)#interface ethernet1/0/1

(config-if-ethernet1/0/1)#mls qos ets enable

(config-if-ethernet1/0/1)#mls qos schedule policy bind p1

Configuration Result:

Enable the function of vlan-shaping in global; the egress of ethernet1/0/1 is bond to p1 of vlan-shaping. When the flow of entering ethernet1/0/1 exceeds the load line and happens to the congestion, the flow of vlan1,2 will enter to level 3 node 1 queue 3 to schedule and the flow of vlan3, 4 will enter to level 3 node 1 queue 4 to schedule. The minimum bandwidth of queue3, 4 which participate the vlan-shaping scheduling is 64kbs and the maximum is 256kbs. The minimum bandwidth of queue3 is 10kbs and the maximum is 200kbs. Queue3, 4 do the schedule according to the wrr algorithm and the proportion of the out flow is 5:2. For the packet whose drop priority is 0 in queue3, it begins to drop packets when the queue length is up to 20% of the maximum length. It begins to drop packets in full speed when the queue length is up to 40% of the maximum length, but the drop packets account for the proportion of all packets to be 80%.

Case2:

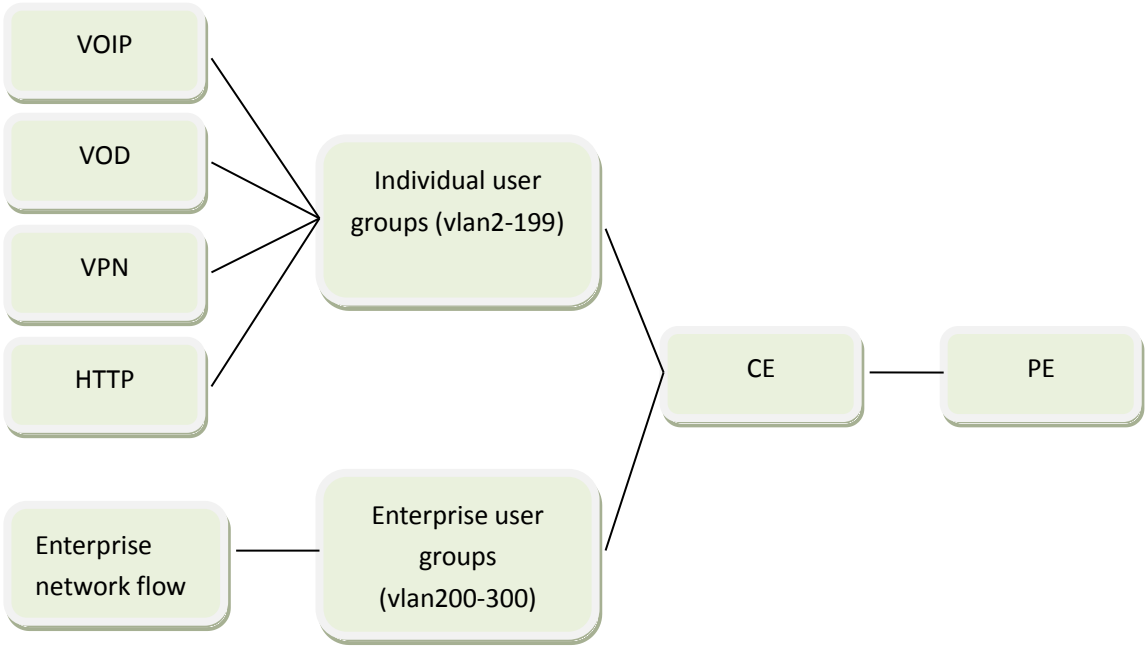


Figure 6-54

In the group network of service-vlan, the users who access the WAN can be divided to individual user and enterprise user. The individual users' business types are consistent and the flow management requirements are consistent too. The enterprise users' business types are different and it requires distinguishing that different enterprise users do different flow management.

Items	Requirement content
the flow policy of individual user groups	The range of service vlan id which carries individual user groups of A is 100 to 299; The individual business can be divided to four kinds: ①VoIP business: The priority of 802.1p is 6 and 7, the minimum bandwidth of the flow whose priority is 6 is 20M, and it is 30M for 7 ②Vod business: The priority of 802.1p is 4 and 5. ③VPN business: The priority of 802.1p is 2 and 3. ④http browsing business: the priority of 802.1p is 0 and 1. ⑤Schedule VoIP business in priority. For the enterprise network flow, Vod, VPN and http business, they can be scheduled according to the proportion of 4:3:2:1.
the flow policy of enterprise user groups	The range of service vlan id which carries enterprise user groups is 500 to 506; Because there are two groups of enterprise users, the two groups' different

	service vlan carry different enterprise users. ①Enterprise users of group A: The range of service vlan id is 501 to 503. ② Enterprise users of group B: The range of service vlan id is 504 to 506. ③ The whole flow requirement of enterprise user groups: The speed limit is 100M; the minimum bandwidth is 50M. ④When scheduling the flow of group A and group B, the proportion is 2:1.

Configuration Thinking:

①Put the enterprise user data into the vlan shaping queues and put the individual user data into the common cos queues. The enterprise users decide the queue number according to vlan id and the individual users decide the queue number according to the priority.

②The individual user data can schedule through S2 and the enterprise user data can schedule through S3.1. the output of S3.1 can be put on S2 for scheduling.

③The queue numbers of VoIP, Vod, VPN and http on S2.2 should be decided according to the priority of 802.1p. So the queue numbers of VoIP are 6 and 7; the queue numbers of Vod are 4 and 5; the queue numbers of VPN are 2 and 3; the queue numbers of http are 0 and 1.

④The flow of VoIP is scheduled for priority and then the flow of Vod, VPN, http and enterprise network are scheduled by the proportion of 3: 2: 1: 4. So put the output of S3.1 and VoIP, VPN, http onto S2.2 to schedule according to WRR. Because S2.2 can support 8 queues at most, the flow of VoIP cannot be put onto S2.2 but can be put onto S2.3 only. Then configure S1 to schedule according to SP, it can ensure that the flow of VoIP can schedule for priority.

⑤There are two groups of enterprise users of A and B. Because there is no differentiation service need of enterprise user in the group, it just need to construct two vlan shaping queues and match them with Vlan501~503 and Vlan504~506 respectively.

⑥A and B do the scheduling as the proportion of 2: 1. So the scheduling algorithm of S3.1 should be configured to be WRR. The queue weight of A is 2 and it is 1 for B.

⑦Because the whole flow requirement of enterprise user groups is 100M for the maximum bandwidth and it is 50M for the minimum bandwidth, it is equivalent to configure the overall minimum and maximum bandwidth of S3.1 as 50M and 100M respectively.

⑧VoIP business has the minimum and maximum bandwidth demand. It is equivalent to configure the minimum and maximum bandwidth for the corresponding queue 6 and 7.

Configuration Steps:

①Configure the minimum and maximum bandwidth demand of VoIP business firstly.

(config)#interface ethernet1/0/1

(config-if-ethernet1/0/1)#mls qos queue 6 bandwidth 20000 0

(config-if-ethernet1/0/1)#mls qos queue 7 bandwidth 0 30000

②Construct two forwarding class and match them with vlan id501～503 and vlan id504～506 respectively.

(config)#mls qos forwarding-class enterpriseA

(config-forwarding-class-enterpriseA)#match s-vid 501-503
(config-forwarding-class-enterpriseA)#exit
(config)#mls qos forwarding-class enterpriseB
(config-forwarding-class-enterpriseB)#match s-vid 504-506
③Create the scheduling policy of p1 and configure the scheduling manner of S1.1 as SP.
(config)#mls qos schedule policy p1
(config-schedule-policy-p1)#mls qos schedule level 1 node 1
(config-schedule-policy-p1-lel1nod1)#mls qos schedule algorithm sp
④Configure the scheduling algorithm of vlan shaping as wrr. The weight of the first vlan shaping queue is 2 and it is 1 for the second queue. Please pay attention that the first and second queues of S3 are not the vlan shaping queues in the switch, from the third, they are vlan shaping queues.
(config-schedule-policy-p1)#mls qos schedule level 3 node 1
(config-schedule-policy-p1-lel3nod1)#mls qos schedule algorithm wrr
(config-schedule-policy-p1-lel3nod1)#mls qos schedule wrr weight 0 0 2 1
⑤Configure the input source of vlan shaping queue. The input source of the first vlan shaping queue is the forwarding class of enterpriseA and the input source of the second is enterpriseB.
(config-schedule-policy-p1-lel3nod1)#mls qos schedule queue 3 input forwarding-class enterpriseA
(config-schedule-policy-p1-lel3nod1)#mls qos schedule queue 4 input forwarding-class enterpriseB
⑥Configure the overall minimum bandwidth of vlan shaping scheduling node as 50M and the maximum is 100M.
(config-schedule-policy-p1-le3nod1)#mls qos bandwidth 50000 100000
⑦Configure the input source of S2.2 as user business and enterprise data. Among of them, the user business is decided the queue numbers according to the priority of 802.1p.
(config-schedule-policy-p1)#mls qos schedule level 2 node 2
(config-schedule-policy-p1-lel2nod2)#mls qos schedule queue 1 input uc 1
(config-schedule-policy-p1-lel2nod2)#mls qos schedule queue 2 input uc 2
(config-schedule-policy-p1-lel2nod2)#mls qos schedule queue 3 input uc 3
(config-schedule-policy-p1-lel2nod2)#mls qos schedule queue 4 input uc 4
(config-schedule-policy-p1-lel2nod2)#mls qos schedule queue 5 input uc 5
(config-schedule-policy-p1-lel2nod2)#mls qos schedule queue 6 input uc 6
(config-schedule-policy-p1-lel2nod2)#mls qos schedule queue 7 input node 1
⑧Configure the scheduling algorithm of S2.2 as wr and configure the corresponding queue weight.
(config-schedule-policy-p1-lel2nod2)#mls qos schedule algorithm wrr
(config-schedule-policy-p1-lel2nod2)#mls qos schedule wrr weight 1 1 2 2 3 3 4
⑨Configure the input source of S2.3 as VoIP.
(config-schedule-policy-p1)#mls qos schedule queue 1 input uc 7
(config-schedule-policy-p1)#mls qos schedule queue 2 input uc 8
⑩Bind the scheduling policy of p1 to the port of 1/0/1
(config)#interface ethernet1/0/1
(config-if-ethernet1/0/1)#mls qos ets enable
(config-if-ethernet1/0/1)#mls qos schedule policy bind p1

6.18.4 Vlan-shaping Troubleshooting

When the effect of the actual flow scheduling does not meet the actual user configuration, the following is the possible problems and the solutions under the premise that the hardware and the cable do not have problems:

1. Make sure that the function of vlan-shaping has been bond to the egress port. The relevant flow will transport according to the relevant queue of vlan-shaping after binding. Otherwise, the general QoS schedule will be conducted according to cos of the flow.
2. Confirm if the flow had been congestion, otherwise, the actual effect may not meet the configuration.
3. Confirm if vlan will be matched to the forwarding class and bond to the schedule policy.
4. The function of vlan shaping is only effective to the unicast packet known; it is not effective to the broadcast packet, multicast packet and the unicast packet unknown.

Chapter 7 Reliability Configuration

7.1 MSTP

7.1.1 Introduction to MSTP

The MSTP (Multiple STP) is a new spanning-tree protocol which is based on the STP and the RSTP. It runs on all the bridges of a bridged-LAN. It calculates a common and internal spanning tree (CIST) for the bridge-LAN which consists of the bridges running the MSTP, the RSTP and the STP. It also calculates the independent multiple spanning-tree instances (MSTI) for each MST domain (MSTP domain). The MSTP, which adopts the RSTP for its rapid convergence of the spanning tree, enables multiple VLANs to be mapped to the same spanning-tree instance which is independent to other spanning-tree instances. The MSTP provides multiple forwarding paths for data traffic and enables load balancing. Moreover, because multiple VLANs share a same MSTI, the MSTP can reduce the number of spanning-tree instances, which consumes less CPU resources and reduces the bandwidth consumption.

7.1.1.1 MSTP Region

Because multiple VLANs can be mapped to a single spanning tree instance, IEEE 802.1s committee raises the MST concept. The MST is used to make the association of a certain VLAN to a certain spanning tree instance.

A MSTP region is composed of one or multiple bridges with the same MCID (MST Configuration Identification) and the bridged-LAN (a certain bridge in the MSTP region is the designated bridge of the LAN, and the bridges attaching to the LAN are not running STP). All the bridges in the same MSTP region have the same MSID.

MSID consists of 3 attributes:

- ☞ Configuration Name: Composed by digits and letters
- ☞ Revision Level
- ☞ Configuration Digest: VLANs mapping to spanning tree instances

The bridges with the same 3 above attributes are considered as in the same MST region.

When the MSTP calculates CIST in a bridged-LAN, a MSTP region is considered as a bridge. See the figure below:

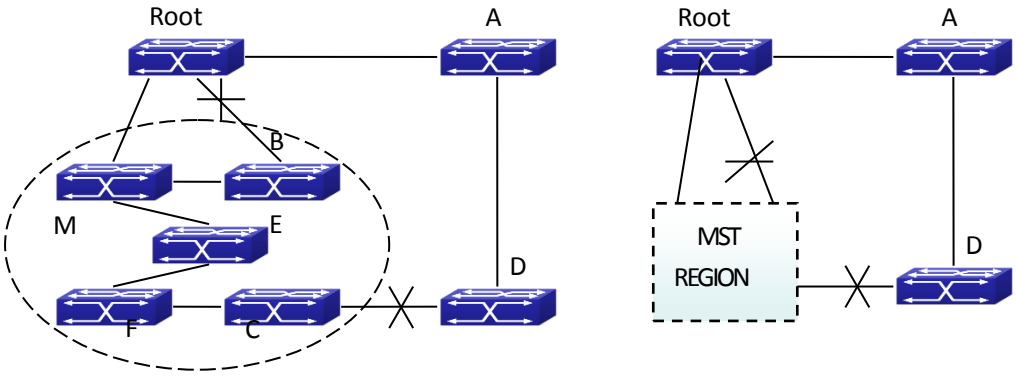


Figure 7-1 Example of CIST and MST Region

In the above network, if the bridges are running the STP or the RSTP, one port between Bridge M and Bridge B should be blocked. But if the bridges in the yellow range run the MSTP and are configured in the same MST region, MSTP will treat this region as a bridge. Therefore, one port between Bridge B and Root is blocked and one port on Bridge D is blocked.

7.1.1.1.1 Operations within an MSTP Region

The IST connects all the MSTP bridges in a region. When the IST converges, the root of the IST becomes the IST master, which is the switch within the region with the lowest bridge ID and path cost to the CST root. The IST master is also the CST root if there is only one region within the network. If the CST root is outside the region, one of the MSTP bridges at the boundary of the region is selected as the IST master.

When an MSTP bridge initializes, it sends BPDUs claiming itself as the root of the CST and the IST master, with both of the path costs to the CST root and to the IST master set to zero. The bridge also initializes all of its MST instances and claims to be the root for all of them. If the bridge receives superior MST root information (lower bridge ID, lower path cost, and so forth) than currently stored for the port, it relinquishes its claim as the IST master.

Within a MST region, the IST is the only spanning-tree instance that sends and receives BPDUs. Because the MST BPDU carries information for all instances, the number of BPDUs that need to be processed by a switch to support multiple spanning-tree instances is significantly reduced.

All MST instances within the same region share the same protocol timers, but each MST instance has its own topology parameters, such as root switch ID, root path cost, and so forth.

7.1.1.1.2 Operations between MST Regions

If there are multiple regions or legacy 802.1D bridges within the network, MSTP establishes and maintains the CST, which includes all MST regions and all legacy STP bridges in the network. The MST instances combine with the IST at the boundary of the region to become the CST.

The MSTI is only valid within its MST region. An MSTI has nothing to do with MSTIs in other MST regions. The bridges in a MST region receive the MST BPDU of other regions through Boundary Ports. They only process CIST related information and abandon MSTI information.

7.1.1.2 Port Roles

The MSTP bridge assigns a port role to each port which runs MSTP.

- ☞ CIST port roles: Root Port, Designated Port, Alternate Port and Backup Port
- ☞ On top of those roles, each MSTI port has one new role: Master Port.

The port roles in the CIST (Root Port, Designated Port, Alternate Port and Backup Port) are defined in the same ways as those in the RSTP.

7.1.1.3 MSTP Load Balance

In a MSTP region, VLANs can by mapped to various instances. That can form various topologies. Each instance is independent from the others and each distance can have its own attributes such as bridge priority and port cost etc. Consequently, the VLANs in different instances have their own paths. The traffic of the VLANs are load-balanced.

7.1.2 MSTP Configuration Task List

MSTP configuration task list:

1. Enable the MSTP and set the running mode
2. Configure instance parameters
3. Configure MSTP region parameters
4. Configure MSTP time parameters
5. Configure the fast migrate feature for MSTP
6. Configure the format of port packet
7. Configure the spanning-tree attribute of port
8. Configure the snooping attribute of authentication key
9. Configure the FLUSH mode once topology changes

1. Enable MSTP and set the running mode

Command	Explanation
Global Mode and Port Mode	
spanning-tree **no spanning-tree**	Enable/Disable MSTP.
Global Mode	
spanning-tree mode {mstp|stp|rstp} **no spanning-tree mode**	Set MSTP running mode.
Port Mode	
spanning-tree mcheck	Force port migrate to run under MSTP.

2. Configure instance parameters

Command	Explanation
Global Mode	
spanning-tree mst *<instance-id>* priority	Set bridge priority for specified instance.

<bridge-priority> **no spanning-tree mst *<instance-id>* priority**	
spanning-tree priority *<bridge-priority>* **no spanning-tree priority**	Configure the spanning-tree priority of the switch.
Port Mode	
spanning-tree mst *<instance-id>* cost *<cost>* **no spanning-tree mst *<instance-id>* cost**	Set port path cost for specified instance.
spanning-tree mst *<instance-id>* port-priority *<port-priority>* **no spanning-tree mst *<instance-id>* port-priority**	Set port priority for specified instance.
spanning-tree mst *<instance-id>* rootguard **no spanning-tree mst *<instance-id>* rootguard**	Configure currently port whether running rootguard in specified instance, configure the rootguard port can't turn to root port.
spanning-tree rootguard **no spanning-tree rootguard**	Configure currently port whether running rootguard in instance 0, configure the rootguard port can't turn to root port.
spanning-tree [mst *<instance-id>*] loopguard **no spanning-tree [mst *<instance-id>*] loopguard**	Enable loopguard function on specified instance, the no command disables this function.

3. Configure MSTP region parameters

Command	Explanation
Global Mode	
spanning-tree mst configuration **no spanning-tree mst configuration**	Enter MSTP region mode. The no command restores the default setting.
MSTP region mode	
show	Display the information of the current running system.
instance *<instance-id>* vlan *<vlan-list>* **no instance *<instance-id>* [vlan *<vlan-list>*]**	Create Instance and set mapping between VLAN and Instance.
name *<name>* **no name**	Set MSTP region name.
revision-level *<level>* **no revision-level**	Set MSTP region revision level.
abort	Quit MSTP region mode and return to Global mode without saving MSTP region configuration.
exit	Quit MSTP region mode and return to Global mode with saving MSTP region configuration.
no	Cancel one command or set initial value.

4. Configure MSTP time parameters

Command	Explanation
Global Mode	
spanning-tree forward-time *<time>* **no spanning-tree forward-time**	Set the value for switch forward delay time.
spanning-tree hello-time *<time>* **no spanning-tree hello-time**	Set the Hello time for sending BPDU messages.
spanning-tree maxage *<time>* **no spanning-tree maxage**	Set Aging time for BPDU messages.
spanning-tree max-hop *<hop-count>* **no spanning-tree max-hop**	Set Maximum number of hops of BPDU messages in the MSTP region.

5. Configure the fast migrate feature for MSTP

Command	Explanation
Port Mode	
spanning-tree link-type p2p {auto|force-true|force-false} **no spanning-tree link-type**	Set the port link type.
spanning-tree portfast [bpdufilter| bpduguard] [recovery <30-3600>] **no spanning-tree portfast**	Set and cancel the port to be an boundary port. bpdufilter receives the BPDU discarding; bpduguard receives the BPDU will disable port; no parameter receives the BPDU, the port becomes a non-boundary port.

6. Configure the format of MSTP

Command	Explanation
Port Mode	
spanning-tree format standard **spanning-tree format privacy** **spanning-tree format auto** **no spanning-tree format**	Configure the format of port spanning-tree packet, standard format is provided by IEEE, privacy is compatible with CISCO and auto means the format is determined by checking the received packet.

7. Configure the spanning-tree attribute of port

Command	Explanation
Port Mode	
spanning-tree cost **no spanning-tree cost**	Set the port path cost.
spanning-tree port-priority **no spanning-tree port-priority**	Set the port priority.

spanning-tree rootguard **no spanning-tree rootguard**	Set the port is root port.
Global Mode	
spanning-tree transmit-hold-count ***<tx-hold-count-value>*** **no spanning-tree transmit-hold-count**	Set the max transmit-hold-count of port.
spanning-tree cost-format {dot1d | dot1t}	Set port cost format with dot1d or dot1t.

8. Configure the snooping attribute of authentication key

Command	Explanation
Port Mode	
spanning-tree digest-snooping **no spanning-tree digest-snooping**	Set the port to use the authentication string of partner port. The no command restores to use the generated string.

9. Configure the FLUSH mode once topology changes

Command	Explanation
Global Mode	
spanning-tree tcflush {enable| disable| protect} **no spanning-tree tcflush**	Enable: the spanning-tree flush once the topology changes. Disable: the spanning tree don't flush when the topology changes. Protect: the spanning-tree flush not more than one time every ten seconds. The no command restores to default setting, enable flush once the topology changes.
Port Mode	
spanning-tree tcflush {enable| disable| protect} **no spanning-tree tcflush**	Configure the port flush mode. The no command restores to use the global configured flush mode.

7.1.3 MSTP Example

The following is a typical MSTP application example:

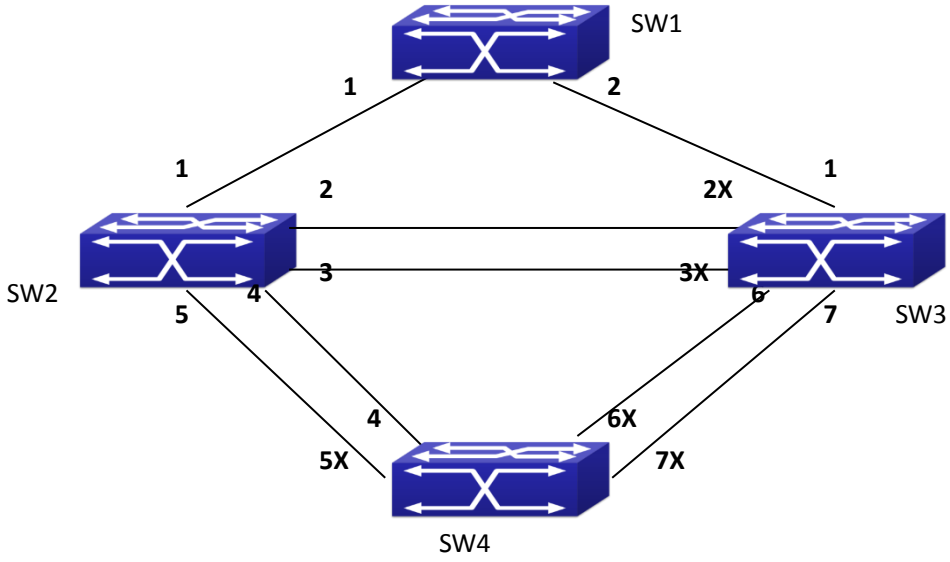


Figure 7-2 Typical MSTP Application Scenario

The connections among the switches are shown in the above figure. All the switches run in the MSTP mode by default, their bridge priority, port priority and port route cost are all in the default values (equal). The default configuration for switches is listed below:

Bridge Name		SW1	SW2	SW3	SW4
Bridge MAC Address		…00-00-01	…00-00-02	…00-00-03	…00-00-04
Bridge Priority		32768	32768	32768	32768
Port Priority	port 1	128	128	128	
	port 2	128	128	128	
	port 3		128	128	
	port 4		128		128
	port 5		128		128
	port 6			128	128
	port 7			128	128
Route Cost	port 1	200000	200000	200000	
	port 2	200000	200000	200000	
	port 3		200000	200000	
	port 4		200000		200000
	port 5		200000		200000
	port 6			200000	200000
	port 7			200000	200000

By default, the MSTP establishes a tree topology (in blue lines) rooted with SwitchA. The ports marked with “x” are in the discarding status, and the other ports are in the forwarding status.

Configurations Steps:

Step 1: Configure port to VLAN mapping:

- ☞ Create VLAN 20, 30, 40, 50 in Switch2, Switch3 and Switch4.
- ☞ Set ports 1-7 as trunk ports in Switch2 Switch3 and Switch4.

Step 2: Set Switch2, Switch3 and Switch4 in the same MSTP:

- ☞ Set Switch2, Switch3 and Switch4 to have the same region name as mstp.
- ☞ Map VLAN 20 and VLAN 30 in Switch2, Switch3 and Switch4 to Instance 3; Map VLAN 40 and VLAN 50 in Switch2, Switch3 and Switch4 to Instance 4.

Step 3: Set Switch3 as the root bridge of Instance 3; Set Switch4 as the root bridge of Instance 4

- ☞ Set the bridge priority of Instance 3 in Switch3 as 0.
- ☞ Set the bridge priority of Instance 4 in Switch4 as 0.

The detailed configuration is listed below:

Switch2:

```
Switch2(config)#vlan 20
Switch2(Config-Vlan20)#exit
Switch2(config)#vlan 30
Switch2(Config-Vlan30)#exit
Switch2(config)#vlan 40
Switch2(Config-Vlan40)#exit
Switch2(config)#vlan 50
Switch2(Config-Vlan50)#exit
Switch2(config)#spanning-tree mst configuration
Switch2(Config-Mstp-Region)#name mstp
Switch2(Config-Mstp-Region)#instance 3 vlan 20;30
Switch2(Config-Mstp-Region)#instance 4 vlan 40;50
Switch2(Config-Mstp-Region)#exit
Switch2(config)#interface e1/0/1-7
Switch2(Config-Port-Range)#switchport mode trunk
Switch2(Config-Port-Range)#exit
Switch2(config)#spanning-tree
```

Switch3:

```
Switch3(config)#vlan 20
Switch3(Config-Vlan20)#exit
Switch3(config)#vlan 30
Switch3(Config-Vlan30)#exit
Switch3(config)#vlan 40
Switch3(Config-Vlan40)#exit
Switch3(config)#vlan 50
Switch3(Config-Vlan50)#exit
```

```
Switch3(config)#spanning-tree mst configuration
Switch3(Config-Mstp-Region)#name mstp
Switch3(Config-Mstp-Region)#instance 3 vlan 20;30
Switch3(Config-Mstp-Region)#instance 4 vlan 40;50
Switch3(Config-Mstp-Region)#exit
Switch3(config)#interface e1/0/1-7
Switch3(Config-Port-Range)#switchport mode trunk
Switch3(Config-Port-Range)#exit
Switch3(config)#spanning-tree
Switch3(config)#spanning-tree mst 3 priority 0
```

Switch4:

```
Switch4(config)#vlan 20
Switch4(Config-Vlan20)#exit
Switch4(config)#vlan 30
Switch4(Config-Vlan30)#exit
Switch4(config)#vlan 40
Switch4(Config-Vlan40)#exit
Switch4(config)#vlan 50
Switch4(Config-Vlan50)#exit
Switch4(config)#spanning-tree mst configuration
Switch4(Config-Mstp-Region)#name mstp
Switch4(Config-Mstp-Region)#instance 3 vlan 20;30
Switch4(Config-Mstp-Region)#instance 4 vlan 40;50
Switch4(Config-Mstp-Region)#exit
Switch4(config)#interface e1/0/1-7
Switch4(Config-Port-Range)#switchport mode trunk
Switch4(Config-Port-Range)#exit
Switch4(config)#spanning-tree
Switch4(config)#spanning-tree mst 4 priority 0
```

After the above configuration, Switch1 is the root bridge of the instance 0 of the entire network. In the MSTP region which Switch2, Switch3 and Switch4 belong to, Switch2 is the region root of the instance 0, Switch3 is the region root of the instance 3 and Switch4 is the region root of the instance 4. The traffic of VLAN 20 and VLAN 30 is sent through the topology of the instance 3. The traffic of VLAN 40 and VLAN 50 is sent through the topology of the instance 4. And the traffic of other VLANs is sent through the topology of the instance 0. The port 1 in Switch2 is the master port of the instance 3 and the instance 4.

The MSTP calculation generates 3 topologies: the instance 0, the instance 3 and the instance 4 (marked with blue lines). The ports with the mark “x” are in the status of discarding. The other ports are the status of forwarding. Because the instance 3 and the instance 4 are only valid in the MSTP region, the following figure only shows the topology of the MSTP region.

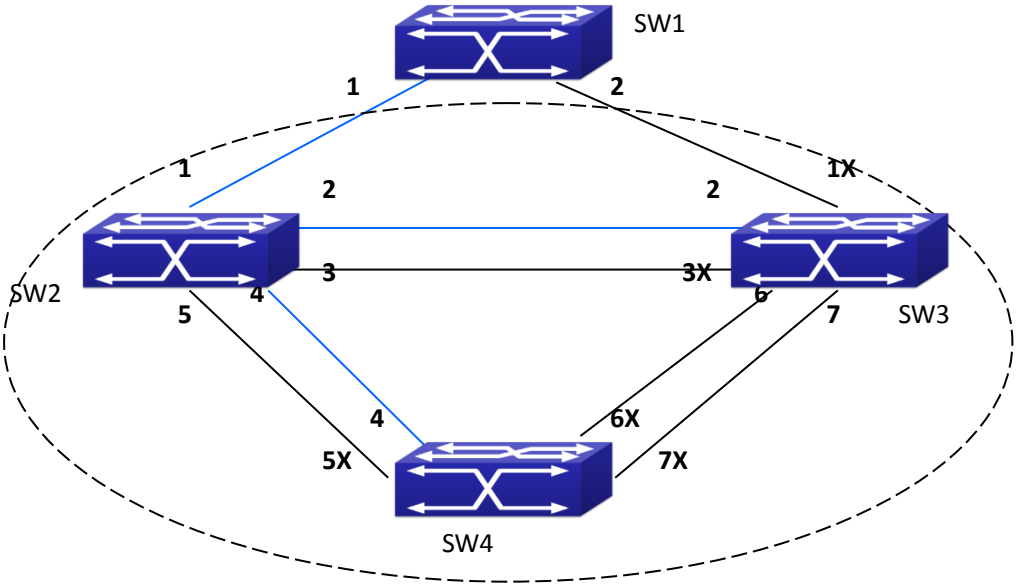


Figure 7-3 The Topology Of the Instance 0 after the MSTP Calculation

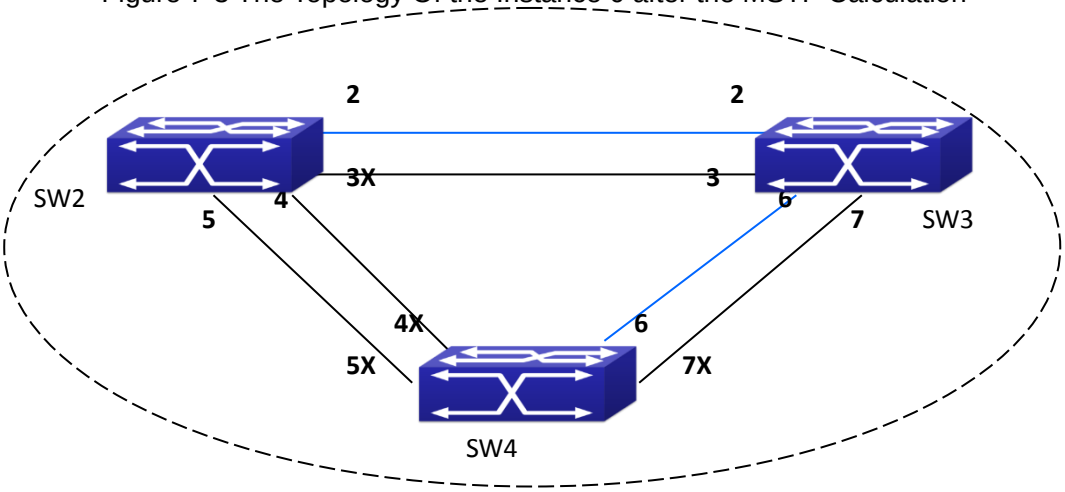


Figure 7-4 The Topology Of the Instance 3 after the MSTP Calculation

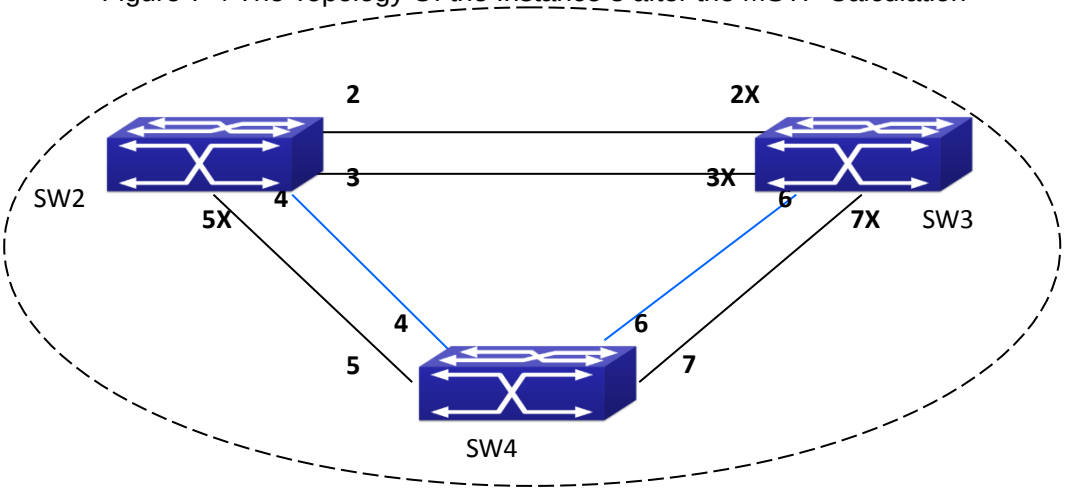


Figure 7-5 The Topology Of the Instance 4 after the MSTP Calculation

7.1.4 MSTP Troubleshooting

- In order to run the MSTP on the switch port, the MSTP has to be enabled globally. If the MSTP is not enabled globally, it can't be enabled on the port.
- The MSTP parameters co work with each other, so the parameters should meet the following conditions. Otherwise, the MSTP may work incorrectly.
  2×(Bridge_Forward_Delay -1.0 seconds) >= Bridge_Max_Age
  Bridge_Max_Age >= 2 ×(Bridge_Hello_Time + 1.0 seconds)
- When users modify the MSTP parameters, they have to be sure about the changes of the topologies. The global configuration is based on the bridge. Other configurations are based on the individual instances.

7.2 ERPS

7.2.1 Introduction to ERPS

ERPS (Ethernet Ring Protection Switching) is a kind of layer2 anti-ring protocol which is defined by ITU-T; the standard number is ITU-T G.8032/Y1344, also known as G.8032. G.8032 Ethernet standard absorbs the advantages of the ring network protection technology of EAPS, RPR, SDH, STP.etc. It optimizes the detection mechanism and it can detect the two-way/single-way fault, support the structure with multi-ring and multi-domain. At the same time of achieving rotating of 50ms, it supports the master and slave and load sharing. It becomes the newest standard of the Ethernet ring technology.

ERPS is the anti-ring protocol used in ring network protection. It includes: link switching in loop fault, notification and loop rotating after loop restored, but it does not include the discovery of link fault. The CCM function defined by 802.1ag protocol can be used in fault discovery and the physical link fault detection can also be used. The principle is that it must be flexible no matter which detection mechanism was used, the link fault can be found in a short time and it will be noticed to the erps module. The link fault rotating time that erps asks is: flow discontinuity time is 50ms at most with that the link length is in 1200km and in 16 nodes. It is demanding for the link fault discovery and loop protection protocol rotating time.

7.2.1.1 ERPS Terminology

Ethernet ring： It is the closed physical ring network which is made up by many ring nodes, every node on the ring has only two ports connecting to this ring network.

Ring protection link： RPL is a link on the ring network. When the ring network is healthy, the link blocked by the node cannot transmit the data flow.

RPL owner node： When the ring network is healthy, the nodes connected to RPL will block the RPL. At the same time, it will launch the link rotating when the ring network restored and it is

configured as reversion.

RPL neighbor node：RPL neighbor node, it is the other node connected to RPL. When the ring network is healthy, it will block the RPL.

Interconnection node：Cross node, when there are many rings are crossed, it is the node in the cross position. On the cross nodes, there is one or more rings can be connected through two ports. The ring connected through one port is the sub-ring, the ring connected through two ports is the main ring.

R-APS virtual channel：It is the link which makes the sub-ring connect between two interconnection nodes out of the sub-ring path. Its transmission characteristic is related to the out ring network.

major-ring： It is the ring which connects the two ports on the interconnection node.

sub-ring：It is the ring which connects to other network through two interconnection nodes. it is not a ring network, it will make up a ring network only when connect it through the interconnection node.

ERP instance：It is a set protected by many vlan. The packet transmission of the vlan in this instance pass the same ring network link, every vlan only belongs to one instance.

Revertive switch：After learning of the ring network fault restored, the RPL owner node will restore the blockade status of RPL and make the network flow transmission path restore to the link before the fault.

Non-revertive switch：After learning of the ring network fault restored, the RPL owner node will not block the RPL, the network flow transmission path is same as before.

7.2.1.2 ERPS Function

7.2.1.2.1 Fault Switchover

The following is the single-ring and single link fault.

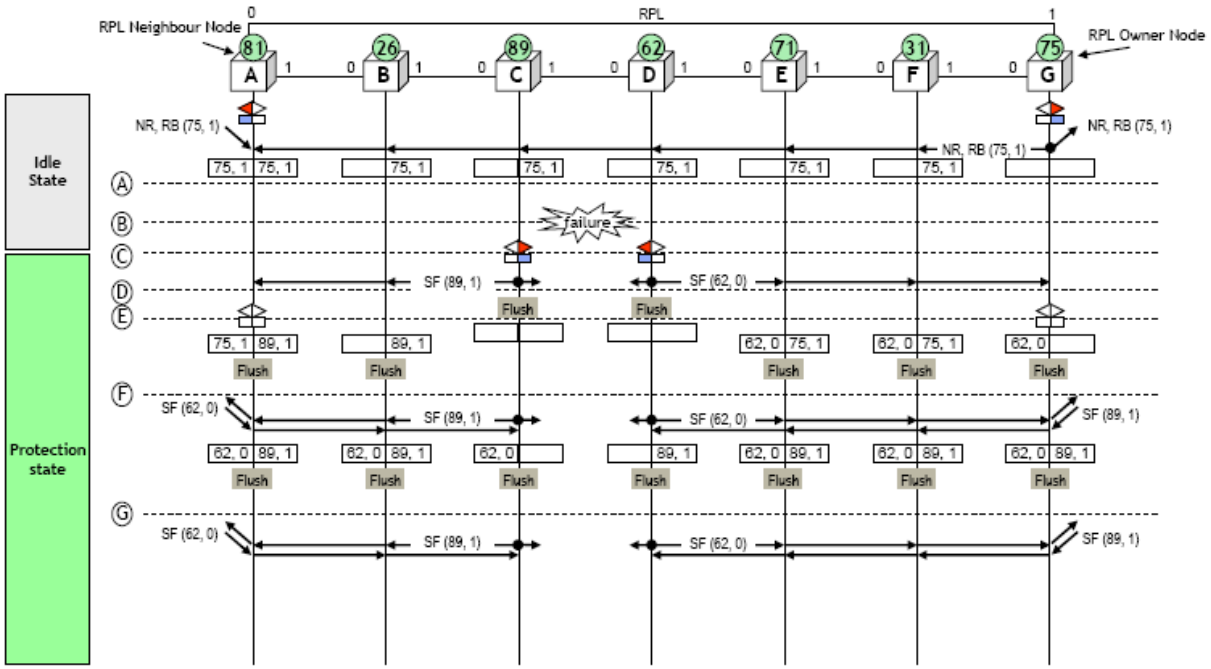


Figure 7-6 single-link fault

The steps of fault switchover:

a) The ring network status is normal; RPL owner node of G sends the R-APS (NR, RB) packet periodically. This packet explains that RPL link is in blockade status and the ring network is healthy.

b) There is fault on the link between node C and D.

c) The node C and D detected the fault, they block the port which connected to the fault link respectively and run the flush FDB.

d) At the same time, the node C and D send the fault notification packet of R-APS (SF) respectively through the port connected to the ring network.

e) All the nodes which received R-APS (SF) packet will run the flush FDB. At the same time, RPL owner node of G and RPL neighbor node of A will configure the RPL connection port as forward. The node G will stop sending R-APS (NR,RB) packet.

f) Because RPL link has removed the blockade, all nodes can receive two R-APS (SF) packets (sent by node C and D). after receiving the new R-APS (SF) packet, it will run the flush FDB.

g) Link fault message of R-APS (SF) will transmitting always in the ring network.

7.2.1.2.2 Failure Recovery

When the ring link restored, there are two methods on the ring nodes: one is Revertive switch. After the ring link restored to be normal, the ring network will block the RPL, and restore the forwarding status of the fault link. At this time, the forwarding path of the data packet is same as the last once when it is normal. Another one is Non-revertive switch. After the ring link restored to be normal, the link will keep the block status. The data packet will continue to be forwarded with the current path.

The environments of the two methods are different. When the block RPL can make the data flow transmission path be the best, use the Revertive switch; when the path costs are similar, there is no difference no matter which path will be blocked, for preventing the secondary interruption of data flow, use the Non-revertive switch.

1. Revertive switch

The following is the single-ring and single link fault.

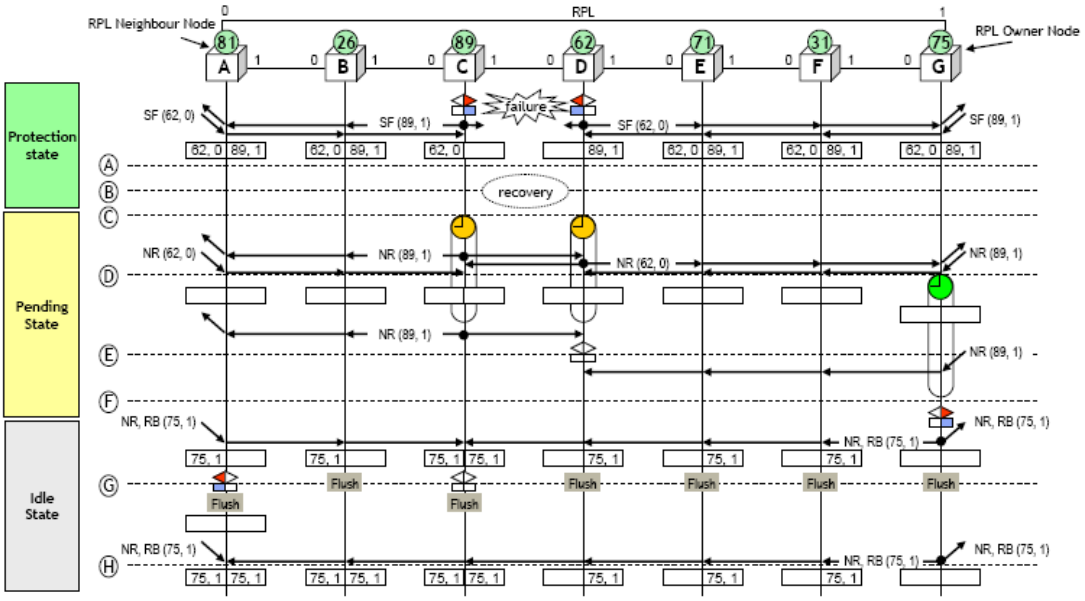


Figure 7-7 Revertive switch fault restoration of single link

The steps:

a) The fault still existed, the node which has detected the fault will send the R-APS (SF) packets with the fault message periodically;

b) Fault restoration on the link;

c) The nodes of C and D detects the fault restoration, they will start the guard timer and send the fault restoration packets of R-APS (NR) on the ports of the ring network at the same time;

d) When the RPL owner node detected R-APS (NR) packets, it will start the WTR timer and clear the local node fault message at the same time;

e) After the nodes C and D are time out, they receives the R-APS (NR) packets from the peer. The node D thinks the priority of node C is higher, so it will stop sending the R-APS (NR) packets with the local message and relieve the block of the port;

f) When the WTR timer of RPL owner node G is time out, it will block the port connected to RPL and send the R-APS (NR, RB) packets through the ring network port to notify other nodes that RPL link has been blocked. At the same time, the node G run the flush FDB;

g) When the node C received the R-APS (NR, RB) packets sent by RPL owner node, it will relieve the block of the local port and stop sending NR packets at the same time. After RPL neighbor node A received this packet, it will block the port

connected to RPL. Other, all nodes will run the flush FDB after received R-APS (NR, RB) packets.

2. Non-revertive switch

The following example is about the single-ring and single link fault as shown in Fig 2-3.

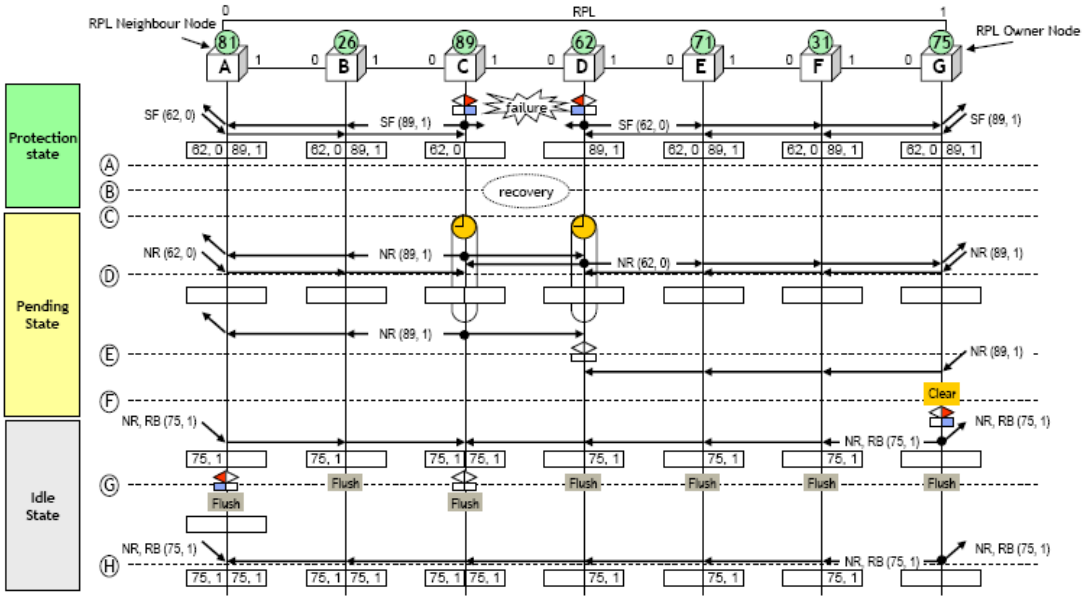


Figure 7-8 Non-revertive switch fault restoration of single link

The steps:

a) The fault still existed, the node which has detected the fault will send the R-APS (SF) packets with the fault message periodically;

b) Fault restoration on the link;

c) The nodes of C and D detects the fault restoration, they will start the guard timer and send the fault restoration packets of R-APS (NR) on the ports of the ring network at the same time;

d) When the RPL owner node G detected the R-APS (NR) packets, it will clear the local node fault message because of the configured non-revertive method, but it will not start the WTR timer;

e) After the nodes C and D are time out, they receives the R-APS (NR) packets from the peer. The node D thinks the priority of node C is higher, so it will stop sending the R-APS (NR) packets and relieve the block of the local port;

f) If RPL owner node G runs clear command, it will be recovered to be revertive method and it will block the port connected to RPL and send R-APS (NR, RB) packets through the ring network port to notify other nodes that the RPL link has been blocked. At the same time, it will run flush FDB;

g) When the node C received the packets sent by RPL owner node, it will relieve the block of the local port and stop sending R-APS (NR) packets at the same time. After RPL neighbor node A received this packet, it will block the port connected to RPL. Other, all nodes will run the flush FDB after received the packets.

7.2.1.2.3 Interconnection Ring Model

ERPS protocol can support the protection and switching of the interconnection ring. The interconnection ring includes two types: the interconnection ring model with virtual channel and the interconnection ring model without virtual channel.

1. The interconnection ring model with virtual channel

As shown in Fig 2-4, three ring networks are interconnection. Ring 1 is the major ring and it is made up with the ring nodes A, B, G, H and the links of them. When ring 1 is health, it will block the link between nodes A and B. ring 2 is another major ring, it is made up with the ring nodes C, D, E, F and the links of them. When ring 2 is health, it will block the link between node C and D. ring 3 is a sub ring, it is made up with the nodes B, C, F, G and the links of B-C and G-F. When ring 3 is health, it will block the link of B-C. B-G links are the interconnection links of ring 1 and ring 3 and it belongs to ring 1. C-F links are the interconnection links of ring 2 and ring 3 and it belongs to ring 2. Ring 1 and ring 2 are both the close ring network, ring 3 is not a ring network. If treat ring 1 as the link between the interconnection nodes B and G of ring 3 (virtual channel), and treat ring 2 as the link between the interconnection nodes C and F of ring 3 (virtual channel), ring 3 will be a ring network.

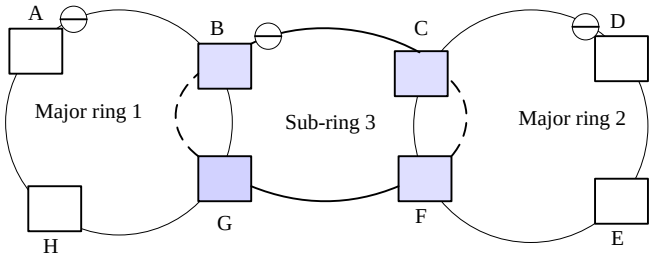


Figure 7-9 the interconnection ring topology with virtual channel

The R-APS virtual channel supported by ring 1 and ring 2 treats ring 3 protocol packets as the data packets. The transmission method of the packets is same as the method of data packets. The node B of ring 3 sends and receives the ring 3 erps protocol packets sent by node G, at the same time, the node G sends and receives the ring 3 erps protocol packets sent by node B. For distinguishing the erps packets of ring 3 and the erps packets of this major ring in ring 1 and ring 2, different control vlan can be used to the protocol packets transmission of every ring.

When the sub ring 3 is changing, it should notify ring 1 and ring 2. The node on the major ring will run flush FDB. The topology changing of the major ring 1 and ring 2 will not affect the sub ring 3. Other, the topology changing of the major ring 1 and ring 2 will not affect each other either.

2. The interconnection ring model without virtual channel

Change the way to understand the ring network as shown in Fig 2-5: ring 1 is the major ring and it is made up with the ring nodes A, B, G, H and the links of them. When ring 1 is health, it will block the link between nodes A and B. ring 2 is sub ring, it is made up with the ring nodes C, D, E, F and the links of C-D, D-E and E-F. When ring 2 is health,

it will block the link between node C and D. ring 3 is another sub ring, it is made up with the nodes B, C, F, G and the links of B-C, C-F and F-G. When ring 3 is health, it will block the link of B-C. B-G links are the interconnection links of ring 1 and ring 3 and it belongs to ring 1. C-F links are the interconnection links of ring 2 and ring 3 and it belongs to ring 3. Ring 1 is the close ring network; ring 2 and ring 3 are not the ring network.

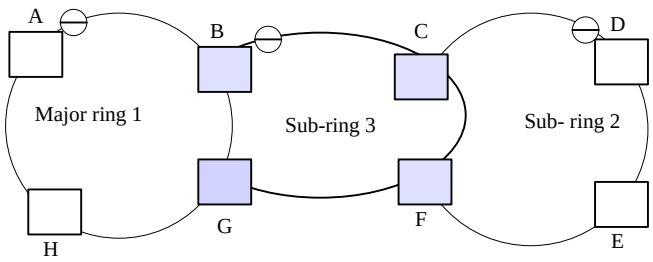


Figure 7-10 the interconnection ring topology without virtual channel

Although ring 2 and ring 3 are not ring network neither, the erps packets of these two sub ring need to be transmitted to all ring nodes. So, the block link should still transmit erps protocol packets if the links of B-C and C-D are blocked, and the nodes B, C and C, D of the block link should also receive and send the erps protocol packets.

When the sub ring 3 is changing, it should notify ring 1. The node on the major ring will run flush FDB. The topology changing will not affect the sub ring 2. When the sub ring 2 is changing, it will affect the sub ring 3 and the major ring 1. The node on the major ring should run flush FDB. But, the topology changing of the major ring 1 will not affect the sub ring 2 and ring 3.

7.2.1.3 ERPS Application

ERPS is used for ring network and it is located in convergence layer, the convergence loop can complete the layer2 convergence of business; insert the layer3 network to deal with the business at the same time. The convergence loop runs ERPS protocol and provides layer2 redundancy protection exchange function of convergence loop.

7.2.2 ERPS Configuration

ERPS Configuration task list as below:

1) Create the instance; the map is corresponding to the vlan which should be protected

2) Create ERPS loop, and configure the member port information. The default configuration: support version V2, the main loop closing type and monitor the physical status of port

3) Configure ERPS loop instance and configure the protection instance, port roles. Configure the ERPS loop instance name, R-APS level, timer information. Configure the controlling VLAN at last and select the port to configure it as RPL owner and RPL Neighbor

1. Create the MSTP instance

Command	Explanation
Global Mode	
spanning-tree mst configuration **no spanning-tree mst configuration**	Enter into the MST configuration mode, configure the parameters of MSTP domain; the no command recovers to be the default.
MST Mode	
instance <instance value > vlan <vlan-list> **no instance [instance-value]**	Configure vlan which needs to be protected by the instance and mapping; the no command deletes the appointed instance.

2. Create ERPS ring and configure the member ports information

Command	Explanation
Global Mode	
erps ring <ring-name> **no erps ring <ring-name>**	Create ERPS ring and enter into the ERPS ring configuration mode; the no command deletes the appointed erps ring.
Port Mode	
erps-ring <ring-name> port0 **erps-ring <ring-name> port1** **erps-ring <ring-name> port0** **erps-ring <ring-name> port1**	Configure the port0 or port1 which is the ring node of port; the no command deletes their property.

3. Configure ERPS ring instance

Command	Explanation
Global Mode	
erps ring <ring-name> **no erps ring <ring-name>**	Create ERPS ring and enter into the ERPS ring configuration mode; the no command deletes the appointed erps ring.
ERPS Ring Configuration Mode	
eprs-instance <instance-id> **no eprs-instance <instance-id>**	Create ERPS ring instance and enter into the ERPS ring configuration mode; the no command deletes the appointed ring node instance.
description *<instance-name>* ***no* description**	Configure the description string of ERPS instance; the no command deletes the appointed string.
rpl {port0 | port1} {owner | neighbour} **no rpl {port0 | port1}**	Configure the member port of ERPS ring instance as RPL owner or RPL neighbor; the no command deletes the appointed owner or neighbor node.
raps-mel <level-value>	Configure the level of R-APS channel, the

no raps-mel	MEL field in the protocol packets is used to detect if the current packet can pass by; the no command deletes the level of R-APS channel.
protected-instance <*instance-list*> **no protected-instance**	Configure the protection instance of ERPS ring instance. The no command deletes the protection instance.
wtr-timer <*wtr-times*> **no wtr-timer**	Configure the WTR timer. The WTR timer is used to avoid the configuration of frequent switching of RPL owner node because of the periodic (discontinuity) fault. The no command deletes the wtr timer.
guard-timer <*guard-times*> **no guard-timer**	Configure the Guard timer. The Guard timer is used in Ethernet ring node to avoid the wrong configuration according to the outdated R-APS packets and avoid the close loop. The no command deletes the guard timer.
holdoff –timer <*holdoff-times*> **no holdoff –timer**	Configure the Holdoff timer. The Holdoff timer is used for Ethernet ring node blocking fault report time. The no command deletes the Holdoff timer.
control-vlan <*vlan-id*> **no control-vlan**	Configure the control vlan of R-APS channel transmission R-APS packets. In the ERPS ring instance, this vlan is used to deliver the ERPS protocol packets but not to forward the user business packets. It improves the security of ERPS protocol. The no command deletes the Control Vlan.

4. Show the configuration information of ERPS

Command	Explanation
Global Mode	
show erps ring {<ring-name> | brief}	Show ERPS ring information.
show erps instance [ring <ring-name> [instance <instance-id>]]	Show ERPS ring instance information.
show erps status [ring <ring-name> [instance <instance-id>]]	Show ERPS ring instance status information.
show erps statistics [ring <ring-name> [instance <instance-id>]]	Show ERPS ring instance statistic information.

7.2.3 ERPS Examples

Case 1:

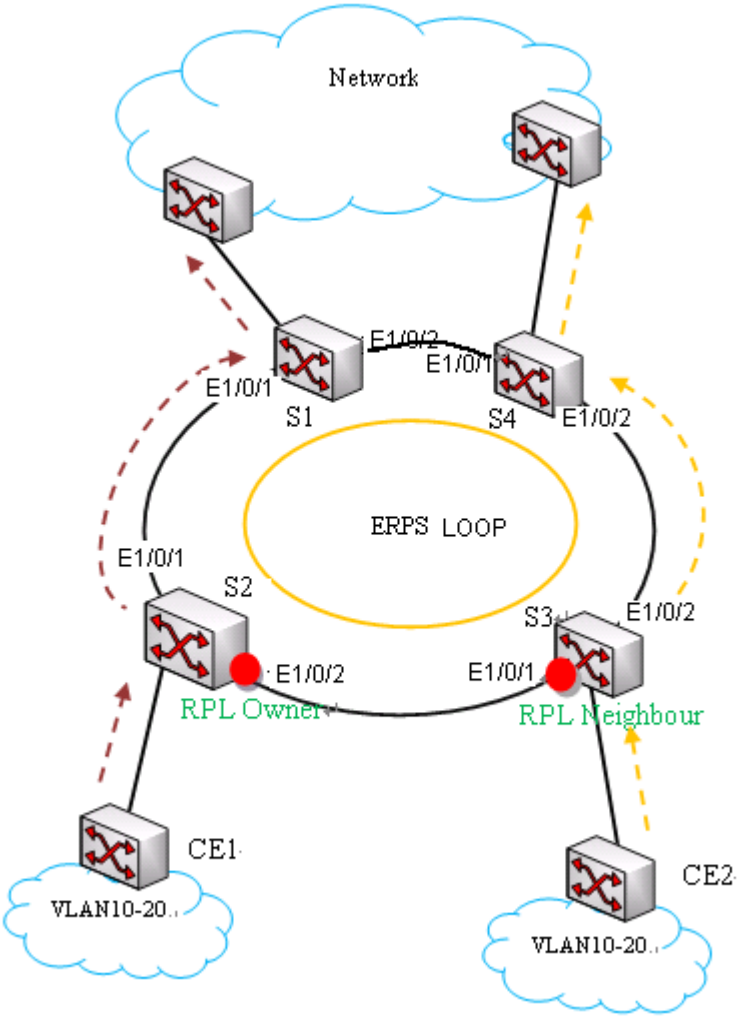


Figure 7-111

As shown in the picture above, it is the explanation of ERPS configuration and application. S1~S4 make up the ring network and provide the layer2 redundancy protection transform function. For preventing the packets in VLAN10 ~ VLAN20 come to be loop, deploy the ERPS protocol on the devices which make up the ring network. The forwarding path of user data inserted through CE1 is S2-S1 and it is S3-S4 for the data which is deserted through CE2. For protecting the Ethernet loop switching, configure it as below:

1. Configuration thinking

Configure ERPS loop redundancy protection as below:

Create ERPS loop of maijor_ring1 and configure the loop member port;

Configure the instance 1 on ERPS loop of maijor_ring1 and configure the protection instance, member port role, timer and controlling VLAN.

2. Configuration steps

Step1: Create instance 2, VLAN2 and VLAN10-20 on S1 ~ S4, VLAN2 is used to transmit the protocol packets, VLAN10-20 are used to transmit the data packets.

Configuration of S1:

S1#config
S1(config)#spanning-tree mst configuration
S1(Config-Mstp-Region) instance 2 vlan 2;10-20
S1(Config-Mstp-Region)#exit
S1(config)#interface e1/1-2
S1(Config-If-Port-Range)#switchport mode trunk
The configuration of S2, S3 and S4 is same as S1.

Step2: Create ERPS loop and configure the member port information. The default configuration: support version V2, main loop closing type and monitor the physical status of port.
Configuration of S1:
S1(config)#erps-ring maijor_ring1
S1(config-erps-ring)#exit
S1(config)# interface e1/1
S1(config-if-ethernet1/1)erps-ring maijor_ring1 port 0
S1(config-if-ethernet1/1)interface e1/2
S1(config-if-ethernet1/2)erps-ring maijor_ring1 port 1

Step3: Configure ERPS loop instance and configure the protection instance, port role. Configure the ERPS loop instance name, R-APS level, timer information. Configure the controlling VLAN at last and configure the port e1/2 of S2 as RPL owner and RPL Neighbor is for e1/1 of S3.
Configuration of S1:
S1(config)# erps-ring maijor_ring1
S1(config-erps-ring)#erps-instance 1
S1(config-erps-ring-inst-1)#description instance1
S1(config-erps-ring-inst-1)#raps-mel 3
S1(config-erps-ring-inst-1)#protected-instance 2
S1(config-erps-ring-inst-1)#wtr-timer 8
S1(config-erps-ring-inst-1)#guard-timer 100
S1(config-erps-ring-inst-1)#holdoff-timer 5
S1(config-erps-ring-inst-1)# control-vlan 2
The configuration of S4 is same as S1.

Configuration of S2:
S2(config)# erps-ring maijor_ring1
S2(config-erps-ring)#erps-instance 1
S2(config-erps-ring-inst-1)#description instance1
S2(config-erps-ring-inst-1)#rpl port 1 owner
S2(config-erps-ring-inst-1)#non-revertive
S2(config-erps-ring-inst-1)#raps-mel 3
S2(config-erps-ring-inst-1)#protected-instance 2
S2(config-erps-ring-inst-1)#wtr-timer 8
S2(config-erps-ring-inst-1)#guard-timer 100
S2(config-erps-ring-inst-1)#holdoff-timer 5

S2(config-erps-ring-inst-1)# control-vlan 2

Configuration of S3:
S3(config)# erps-ring maijor_ring1
S3(config-erps-ring)#erps-instance 1
S3(config-erps-ring-inst-1)#description instance1
S3(config-erps-ring-inst-1)# rp0 port 1 neighbour
S3(config-erps-ring-inst-1)#raps-mel 3
S3(config-erps-ring-inst-1)#protected-instance 2
S3(config-erps-ring-inst-1)#wtr-timer 8
S3(config-erps-ring-inst-1)#guard-timer 100
S3(config-erps-ring-inst-1)#holdoff-timer 5
S3(config-erps-ring-inst-1)# control-vlan 2

Step 4: Check the configuration result. After the configuration above is successful, check the configuration result and below is for S2.

S2# show erps ring brief

Ring-ID	Description	Ring-topo	Port0	Port1	Version	Inst-Count
1	maijor_ring1	maijor-ring	1/1	1/2	V2	1

Switch#show erps instance
ERPS Ring maijor_ring1
Instance 1
Description：instance1
Protected Instance : 2 Revertive mode: revertive
RAPS MEL: 3 R-APS-Virtual-Channel：
Control Vlan : 2
Guard Timer (csec) : 100
Holdoff Timer (seconds) : 5
WTR Timer (min) : 8

Port	Role	Port-Status
port0	Common	Forwarding
port1	RPL Owner	Blocked

3. Configure the file
The configuration file of S1:
S1#config
S1(config)#erps-ring maijor_ring1
S1(config)#spanning-tree mst configuration
S1(Config-Mstp-Region) instance 2 vlan 2;10-20
S1(Config-Mstp-Region)#exit
S1(config)#interface e1/1-2
S1(Config-If-Port-Range)#switchport mode trunk

```
S1(Config-If-Port-Range)#exit
S1(config)# interface e1/1
S1(config-if-ethernet1/1)erps-ring maijor_ring1 port 0
S1(config-if-ethernet1/1)interface e1/2
S1(config-if-ethernet1/2)erps-ring maijor_ring1 port 1
S1(config-if-ethernet1/2)exit
S1(config)#erps-ring maijor_ring1
S1(config-erps-ring)#erps-instance 1
S1(config-erps-ring-inst-1)#description instance1
S1(config-erps-ring-inst-1)#raps-mel 3
S1(config-erps-ring-inst-1)#protected-instance 2
S1(config-erps-ring-inst-1)#wtr-timer 8
S1(config-erps-ring-inst-1)#guard-timer 100
S1(config-erps-ring-inst-1)#holdoff-timer 5
S1(config-erps-ring-inst-1)# control-vlan 2
```

The configuration file of S2:

```
S2#config
S2(config)#erps-ring maijor_ring1
S2(config)#spanning-tree mst configuration
S2(Config-Mstp-Region) instance 2 vlan 2;10-20
S2(Config-Mstp-Region)#exit
S2(config)#interface e1/1-2
S2(Config-If-Port-Range)#switchport mode trunk
S2(Config-If-Port-Range)#exit
S2(config)# interface e1/1
S2(config-if-ethernet1/1)erps-ring maijor_ring1 port 0
S2(config-if-ethernet1/1)interface e1/2
S2(config-if-ethernet1/2)erps-ring maijor_ring1 port 1
S2(config-if-ethernet1/2)exit
S2(config)#erps-ring maijor_ring1
S2(config-erps-ring)#erps-instance 1
S2(config-erps-ring-inst-1)#description instance1
S2(config-erps-ring-inst-1)#rpl port1 owner
S2(config-erps-ring-inst-1)#non-revertive
S2(config-erps-ring-inst-1)#raps-mel 3
S2(config-erps-ring-inst-1)#protected-instance 2
S2(config-erps-ring-inst-1)#wtr-timer 8
S2(config-erps-ring-inst-1)#guard-timer 100
S2(config-erps-ring-inst-1)#holdoff-timer 5
S2(config-erps-ring-inst-1)# control-vlan 2
```

The configuration file of S3:

```
S3#config
```

S3(config)#erps-ring maijor_ring1
S3(config)#spanning-tree mst configuration
S3(Config-Mstp-Region) instance 2 vlan 2;10-20
S3(Config-Mstp-Region)#exit
S3(config)#interface e1/1-2
S3(Config-If-Port-Range)#switchport mode trunk
S3(Config-If-Port-Range)#exit
S3(config)# interface e1/1
S3(config-if-ethernet1/1)erps-ring maijor_ring1 port 0
S3(config-if-ethernet1/1)interface e1/2
S3(config-if-ethernet1/2)erps-ring maijor_ring1 port 1
S3(config-if-ethernet1/2)exit
S3(config)#erps-ring maijor_ring1
S3(config-erps-ring)#erps-instance 1
S3(config-erps-ring-inst-1)#description instance1
S3(config-erps-ring-inst-1)#rpl port1 neighbour
S3(config-erps-ring-inst-1)#raps-mel 3
S3(config-erps-ring-inst-1)#protected-instance 2
S3(config-erps-ring-inst-1)#wtr-timer 8
S3(config-erps-ring-inst-1)#guard-timer 100
S3(config-erps-ring-inst-1)#holdoff-timer 5
S3(config-erps-ring-inst-1)# control-vlan 2

The configuration file of S4:
S4#config
S4(config)#erps-ring maijor_ring1
S4(config)#spanning-tree mst configuration
S4(Config-Mstp-Region) instance 2 vlan 2;10-20
S4(Config-Mstp-Region)#exit
S4(config)#interface e1/1-2
S4(Config-If-Port-Range)#switchport mode trunk
S4(Config-If-Port-Range)#exit
S4(config)# interface e1/1
S4(config-if-ethernet1/1)erps-ring maijor_ring1 port 0
S4(config-if-ethernet1/1)interface e1/2
S4(config-if-ethernet1/2)erps-ring maijor_ring1 port 1
S4(config-if-ethernet1/2)exit
S4(config)#erps-ring maijor_ring1
S4(config-erps-ring)#erps-instance 1
S4(config-erps-ring-inst-1)#description instance1
S4(config-erps-ring-inst-1)#raps-mel 3
S4(config-erps-ring-inst-1)#protected-instance 2
S4(config-erps-ring-inst-1)#wtr-timer 8
S4(config-erps-ring-inst-1)#guard-timer 100

S4(config-erps-ring-inst-1)#holdoff-timer 5
S4(config-erps-ring-inst-1)# control-vlan 2

7.2.4 ERPS Troubleshooting

If the configured ERPS loop cannot achieve the Ethernet loop switching protection, check if it was wrong with the following reasons:

- ☞ Check if the basic configuration is correct and check if the protection instance of every node, control-vlan, wtr-timer, guard-timer and raps-mel are consistent.
- ☞ Check if the vlan that user data flow is in is not the same one that control-vlan is in. In the ERPS loop instance, control vlan is only used to transmit ERPS protocol packet but not the user business packet; it improves the security of ERPS protocol. User ensures the uniqueness of the configuration. This VLAN is as the vlan tag when sending R-APS packet. In the instance, the protection VLAN configuration of all nodes must be consistent.
- ☞ We suggest the port which user configured on ERPS node is trunk port and ensure that the vlan and control vlan that data packet is in are in the protection instance and ERPS only protect the data and protocol packet in the instance. For instance, the switch enables a protocol (CFM, EFM, Layer3 interface) and it makes the switch send the protocol packet out. Then, if the vlan ID which sends the packet is not in the protection instance, there will be the loop in the topology.
- ☞ If configure the port status test method as fastlink, the hardware must support fastlink function. Break off the notification of port status changing; disable the mac soft study function at the same time.
- ☞ If it is configured associating with CFM, the hardware must support CC function and it can achieve the ability of CCM sending packet in 3.3ms.

7.3 VRRP

7.3.1 Introduction to VRRP

VRRP (Virtual Router Redundancy Protocol) is a fault tolerant protocol designed to enhance connection reliability between routers (or L3 Ethernet switches) and external devices. It is developed by the IETF for local area networks (LAN) with multicast/broadcast capability (Ethernet is a Configuration Example) and has wide applications.

All hosts in one LAN generally have a default route configured to specified default gateway, any packet destined to an address outside the native segment will be sent to the default gateway via this default route. These hosts in the LAN can communicate with the external networks.

However, if the communication link connecting the router serving as default game and external networks fails, all hosts using that gateway as the default next hop route will be unable to communicate with the external networks.

VRRP emerged to resolve such problem. VRRP runs on multiple routers in a LAN, simulating a "virtual" router (also referred to as a "Standby cluster") with the multiple routes. There is an active router (the "Master") and one or more backup routers (the "Backup") in the Standby cluster. The workload of the virtual router is actually undertaken by the active router, while the Backup routers serve as backups for the active router.

The virtual router has its own "virtual" IP address (can be identical with the IP address of some router in the Standby cluster), and routers in the Standby cluster also have their own IP address. Since VRRP runs on routes or Ethernet Switches only, the Standby cluster is transparent to the hosts with the segment. To them, there exists only the IP address of the Virtual Router instead of the actual IP addresses of the Master and Backup(s). And the default gateway setting of all the hosts uses the IP address of the Virtual Router. Therefore, hosts within the LAN communicate with the other networks via this Virtual Router. But basically, they are communicating with the other networks via the Master. In the case when the Master of the Standby cluster fails, a backup will take over its task and become the Master to serve all the hosts in the LAN, so that uninterrupted communication between LAN hosts and external networks can be achieved.

To sum it up, in a VRRP Standby cluster, there is always a router/Ethernet serving as the active router (Master), while the rest of the Standby cluster servers act as the backup router(s) (Backup, can be multiple) and monitor the activity of Master all the time. Should the Master fail, a new Master will be elected by all the Backups to take over the work and continue serving the hosts within the segment. Since the election and take-over duration is brief and smooth, hosts within the segment can use the Virtual Router as normal and uninterrupted communication can be achieved.

7.3.2 VRRP Configuration Task List

Configuration Task List:

1. Create/Remove the Virtual Router (required)
2. Configure VRRP dummy IP and interface (required)
3. Activate/Deactivate Virtual Router (required)
4. Configure VRRP sub-parameters (optional)
   （1） Configure the preemptive mode for VRRP
   （2） Configure VRRP priority
   （3） Configure VRRP Timer intervals
   （4） Configure VRRP interface monitor
   （5） Configure the status of the VRRP session monitor port

1. Create/Remove the Virtual Router

Command	Explanation
Global Mode	

router vrrp <*vrid*> **no router vrrp <*vrid*>**	Creates/Removes the Virtual Router.

2. Configure VRRP Dummy IP Address and Interface

Command	Explanation
VRRP protocol configuration mode	
virtual-ip <*ip*> **no virtual-ip**	Configures VRRP Dummy IP address; the **"no virtual-ip"** command removes the virtual IP address.
interface {IFNAME | Vlan <*ID*>} **no interface**	Configures VRRP interface, the "**no interface**" command removes the interface.

3. Activate/Deactivate Virtual Router

Command	Explanation
VRRP protocol configuration mode	
enable	Activates the Virtual Router.
disable	Deactivates the Virtual Router.

4. Configure VRRP Sub-parameters

(1) Configure the preemptive mode for VRRP

Command	Explanation
VRRP protocol configuration mode	
preempt-mode {true| false}	Configures the preemptive mode for VRRP.

(2) Configure VRRP priority

Command	Explanation
VRRP protocol configuration mode	
priority <*priority*>	Configures VRRP priority.

(3) Configure VRRP Timer intervals

Command	Explanation
VRRP protocol configuration mode	
advertisement-interval <*time*>	Configures VRRP timer value (in seconds).

(4) Configure VRRP interface monitor

Command	Explanation
VRRP protocol configuration mode	
circuit-failover {IFNAME | Vlan <*ID*> } <*value_reduced*> **no circuit-failover**	Configures VRRP interface monitor, the **"no circuit-failover"** removes monitor to the interface.

(5) Configure the status of the VRRP session monitor port

Command	Explanation
VRRP protocol configuration mode	

vrrp track interface {ethernet IFNAME|IFNAME} priority <priority_value> **no vrrp track interface {ethernet IFNAME|IFNAME}**	Configure the status of the VRRP session monitor port.

7.3.3 VRRP Typical Examples

As shown in the figure below, SwitchA and SwitchB are Layer three Ethernet Switches in the same group and provide redundancy for each other.

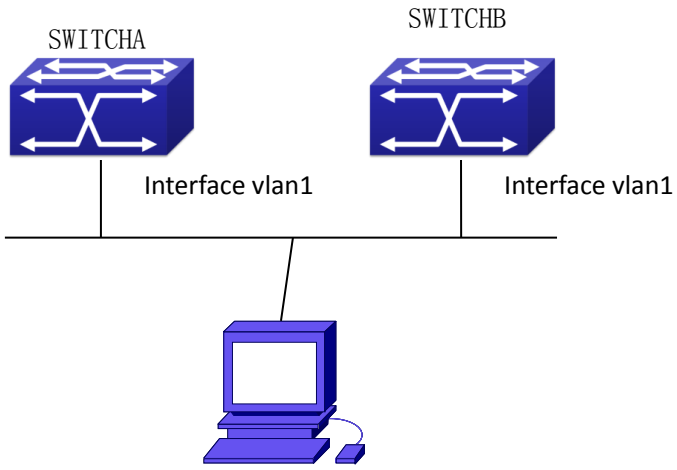


Figure 7-12 VRRP Network Topology

Configuration of SwitchA:
SwitchA(config)#interface vlan 1
SwitchA (Config-if-Vlan1)# ip address 10.1.1.1 255.255.255.0
SwitchA (config)#router vrrp 1
SwitchA(Config-Router-Vrrp)# virtual-ip 10.1.1.5
SwitchA(Config-Router-Vrrp)# interface vlan 1
SwitchA(Config-Router-Vrrp)# enable
Configuration of SwitchB:
SwitchB(config)#interface vlan 1
SwitchB (Config-if-Vlan1)# ip address 10.1.1.7 255.255.255.0
SwitchB(config)#router vrrp 1
SwitchB (Config-Router-Vrrp)# virtual-ip 10.1.1.5
SwitchB(Config-Router-Vrrp)# interface vlan 1
SwitchB(Config-Router-Vrrp)# enable

7.3.4 VRRP Troubleshooting

In configuring and using VRRP protocol, the VRRP protocol may fail to run properly due to reasons such as physical connection failure or wrong configurations. The user should ensure the following:

- ☞ Good condition of the physical connection.
- ☞ All interface and link protocols are in the UP state (use "**show interface**" command).
- ☞ Ensure VRRP is enabled on the interface. Verify the authentication mode of different routers (or L3 Ethernet switches) in the same standby cluster are the same.
- ☞ Verify the timer time of different routers (or L3 Ethernet switches) in the same standby cluster are the same.
- ☞ Verify the dummy IP address is in the same network segment of the interface's actual IP address.
- ☞ If the examination remains unsolved, please use **debug vrrp** and other debugging command and copy the DEBUG message within 3 minutes, send the recorded message to the technical server center of our company.

7.4 MRPP

7.4.1 Introduction to MRPP

MRPP (Multi-layer Ring Protection Protocol), is a link layer protocol applied on Ethernet loop protection. It can avoid broadcast storm caused by data loop on Ethernet ring, and restore communication among every node on ring network when the Ethernet ring has a break link. MRPP is the expansion of EAPS (Ethernet link automatic protection protocol).

MRPP protocol is similar to STP protocol on function, MRPP has below characters, compare to STP protocol:

<1> MRPP specifically uses to Ethernet ring topology

<2> fast convergence, less than 1 s. ideally it can reach 100-50 ms.

7.4.1.1 Conception Introduction

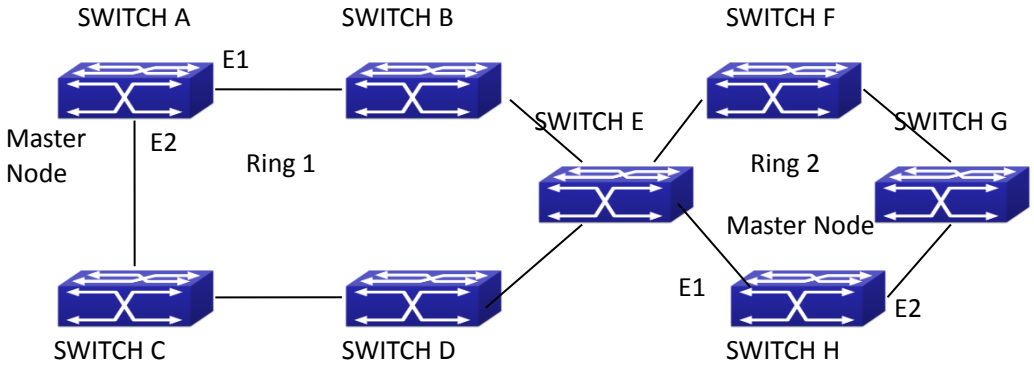


Figure 7-13 MRPP Sketch Map

1. Control VLAN

Control VLAN is a virtual VLAN, only used to identify MRPP protocol packet transferred in the link. To avoid confusion with other configured VLAN, avoids configuring control VLAN ID to be the same with other configured VLAN ID. The different MRPP ring should configure the different control VLAN ID.

2. Ethernet Ring (MRPP Ring)

Ring linked Ethernet network topology.

Each MRPP ring has two states.

Health state: The whole ring net work physical link is connected.

Break state: one or a few physical link break in ring network

3. nodes

Each switch is named after a node on Ethernet. The node has some types:

Primary node: each ring has a primary node, it is main node to detect and defend.

Transfer node: except for primary node, other nodes are transfer nodes on each ring.

The node role is determined by user configuration. As shown Fig 3-1, Switch A is primary node of Ring 1, Switch B. Switch C; Switch D and Switch E are transfer nodes of Ring 1.

4. Primary port and secondary port

The primary node and transfer node have two ports connecting to Ethernet separately, one is primary port, and another is secondary port. The role of port is determined by user configuration.

Primary port and secondary port of primary node.

The primary port of primary node is used to send ring health examine packet (hello), the secondary port is used to receive Hello packet sending from primary node. When the Ethernet is in health state, the secondary port of primary node blocks other data in logical and only MRPP packet can pass. When the Ethernet is in break state, the secondary port of primary node releases block state, and forwards data packets.

There are no difference on function between Primary port and secondary port of transfer

node.

The role of port is determined by user configuration. As shown Fig 3-1, Switch A E1 is primary port, E2 is secondary port.

5. Timer

The two timers are used when the primary node sends and receives MRPP protocol packet: Hello timer and Fail Timer.

Hello timer: define timer of time interval of health examine packet sending by primary node primary port.

Fail timer: define timer of overtime interval of health examine packet receiving by primary node primary port. The value of Fail timer must be more than or equal to the 3 times of value of Hello timer.

7.4.1.2 MRPP Protocol Packet Types

Packet Type	Explanation
Hello packet (Health examine packet) Hello	The primary port of primary node evokes to detect ring, if the secondary port of primary node can receive Hello packet in configured overtime, so the ring is normal.
LINK-DOWN (link Down event packet)	After transfer node detects Down event on port, immediately sends LINK-DOWN packet to primary node, and inform primary node ring to fail.
LINK-DOWN-FLUSH_FDB packet	After primary node detects ring failure or receives LINK-DOWN packet, open blocked secondary port, and then uses two ports to send the packet, to inform each transfer node to refresh own MAC address.
LINK-UP-FLUSH_FDB packet	After primary detects ring failure to restore normal, and uses packet from primary port, and informs each transfer node to refresh own MAC address.

7.4.1.3 MRPP Protocol Operation System

1. Link Down Alarm System

When transfer node finds themselves belonging to MRPP ring port Down, it sends link Down packet to primary node immediately. The primary node receives link down packet and immediately releases block state of secondary port, and sends LINK-DOWN-FLUSH-FDB packet to inform all of transfer nodes, refreshing own MAC address forward list.

2. Poll System

The primary port of primary node sends Hello packet to its neighbors timely according to configured Hello-timer.

If the ring is health, the secondary port of primary node receives health detect packet, and

the primary node keeps secondary port.

If the ring is break, the secondary port of primary node can't receive health detect packet when timer is over time. The primary releases the secondary port block state, and sends LINK-DOWN-FLUSH_FDB packet to inform all of transfer nodes, to refresh own MAC address forward list.

3. Ring Restore

After the primary node occur ring fail, if the secondary port receives Hello packet sending from primary node, the ring has been restored, at the same time the primary node block its secondary port, and sends its neighbor LINK-UP-Flush-FDB packet.

After MRPP ring port refresh UP on transfer node, the primary node maybe find ring restore after a while. For the normal data VLAN, the network maybe forms a temporary ring and creates broadcast storm. To avoid temporary ring, transfer node finds it to connect to ring network port to refresh UP, immediately block temporarily (only permit control VLAN packet pass), after only receiving LINK-UP-FLUSH-FDB packet from primary node, and releases the port block state.

7.4.2 MRPP Configuration Task List

1) Globally enable MRPP
2) Configure MRPP ring
3) Configure the query time of MRPP
4) Configure the compatible mode
5) Display and debug MRPP relevant information

1) Globally enable MRPP

Command	Explanation
Global Mode	
mrpp enable **no mrpp enable**	Globally enable and disable MRPP.

2) Configure MRPP ring

Command	Explanation
Global Mode	
mrpp ring <*ring-id*> **no mrpp ring <*ring-id*>**	Create MRPP ring. The "no" command deletes MRPP ring and its configuration.
MRPP ring mode	
control-vlan <*vid*> **no control-vlan**	Configure control VLAN ID, format "no" deletes configured control VLAN ID.
node-mode {master | transit}	Configure node type of MRPP ring (primary node or secondary node).
hello-timer < *timer*> **no hello-timer**	Configure Hello packet timer sending from primary node of MRPP ring, format "no" restores default timer value.

fail-timer <*timer*> **no fail-timer**	Configure Hello packet overtime timer sending from primary node of MRPP ring, format “no” restores default timer value.
enable **no enable**	Enable MRPP ring, format “no” disables enabled MRPP ring.
Port mode	
mrpp ring <*ring-id*> primary-port **no mrpp ring <ring-id> primary-port**	Specify primary port of MRPP ring.
mrpp ring <*ring-id*> secondary-port **no mrpp ring <ring-id> secondary-port**	Specify secondary port of MRPP ring.

3) Configure the query time of MRPP

Command	Explanation
Global Mode	
mrpp poll-time <20-2000>	Configure the query interval of MRPP.

4) Configure the compatible mode

Command	Explanation
Global Mode	
mrpp errp compatible **no mrpp errp compatible**	Enable the compatible mode for ERRP, the no command disables the compatible mode.
mrpp eaps compatible **no mrpp eaps compatible**	Enable the compatible mode for EAPS, the no command disables the compatible mode.
errp domain <*domain-id*> **no errp domain <*domain-id*>**	Create ERRP domain, the no command deletes the configured ERRP domain.

5) Display and debug MRPP relevant information

Command	Explanation
Admin Mode	
debug mrpp **no debug mrpp**	Disable MRPP module debug information, format “no” disable MRPP debug information output.
show mrpp {<*ring-id*>}	Display MRPP ring configuration information.
show mrpp statistics {<*ring-id*>}	Display receiving data packet statistic information of MRPP ring.
clear mrpp statistics {<*ring-id*>}	Clear receiving data packet statistic information of MRPP ring.

7.4.3 MRPP Typical Scenario

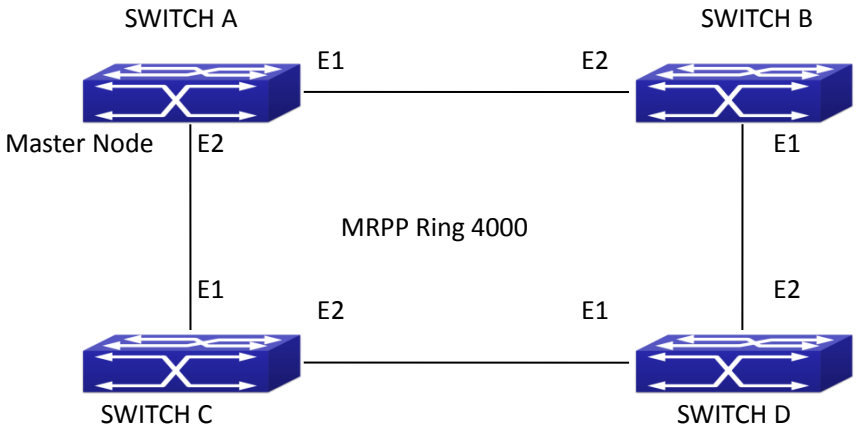


Figure 7-14 MRPP typical configuration scenario

The above topology often occurs on using MRPP protocol. The multi switch constitutes a single MRPP ring, all of the switches only are configured an MRPP ring 4000, thereby constitutes a single MRPP ring.

In above configuration, SWITCH A configuration is primary node of MRPP ring 4000, and configures E1/0/1 to primary port, E1/0/2 to secondary port. Other switches are secondary nodes of MRPP ring, configures primary port and secondary port separately.

To avoid ring, it should temporarily disable one of the ports of primary node, when it enables each MRPP ring in the whole MRPP ring; and after all of the nodes are configured, open the port.

When disable MRPP ring, it needs to insure the MRPP ring doesn't have ring.

SWITCH A configuration Task Sequence:

```
Switch(Config)#mrpp enable
Switch(Config)#mrpp ring 4000
Switch(mrpp-ring-4000)#control-vlan 4000
Switch(mrpp-ring-4000)#fail-timer 18
Switch(mrpp-ring-4000)#hello-timer 5
Switch(mrpp-ring-4000)#node-mode master
Switch(mrpp-ring-4000)#enable
Switch(mrpp-ring-4000)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)#mrpp ring 4000 primary-port
Switch(config-If-Ethernet1/0/1)#interface ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)#mrpp ring 4000 secondary-port
Switch(config-If-Ethernet1/0/2)#exit
Switch(Config)#
```

SWITCH B configuration Task Sequence:

```
Switch(Config)#mrpp enable
Switch(Config)#mrpp ring 4000
Switch(mrpp-ring-4000)#control-vlan 4000
Switch(mrpp-ring-4000)#enable
```

```
Switch(mrpp-ring-4000)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)#mrpp ring 4000 primary-port
Switch(config-If-Ethernet1/0/1)#interface ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)#mrpp ring 4000 secondary-port
Switch(config-If-Ethernet1/0/2)#exit
Switch(Config)#
```

SWITCH C configuration Task Sequence:

```
Switch(Config)#mrpp enable
Switch(Config)#mrpp ring 4000
Switch(mrpp-ring-4000)#control-vlan 4000
Switch(mrpp-ring-4000)#enable
Switch(mrpp-ring-4000)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)#mrpp ring 4000 primary-port
Switch(config-If-Ethernet1/0/1)#interface ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)#mrpp ring 4000 secondary-port
Switch(config-If-Ethernet1/0/2)#exit
Switch(Config)#
```

SWITCH D configuration Task Sequence:

```
Switch(Config)#mrpp enable
Switch(Config)#mrpp ring 4000
Switch(mrpp-ring-4000)#control-vlan 4000
Switch(mrpp-ring-4000)#enable
Switch(mrpp-ring-4000)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)#mrpp ring 4000 primary-port
Switch(config-If-Ethernet1/0/1)#interface ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)#mrpp ring 4000 secondary-port
Switch(config-If-Ethernet1/0/2)#exit
Switch(Config)#
```

7.4.4 MRPP Troubleshooting

The normal operation of MRPP protocol depends on normal configuration of each switch on MRPP ring, otherwise it is very possible to form ring and broadcast storm:

- Configuring MRPP ring, you'd better disconnected the ring, and wait for each switch configuration, then open the ring.
- When the MRPP ring of enabled switch is disabled on MRPP ring, it ensures the ring of the MRPP ring has been disconnected.
- When there is broadcast storm on MRPP ring, it disconnects the ring firstly, and ensures if

each switch MRPP ring configuration on the ring is correct or not; if correct, restores the ring, and then observes the ring is normal or not.

- ☞ The convergence time of MRPP ring net is relative to the response mode of up/down. If use poll mode, the convergence time as hundreds of milliseconds in simple ring net, if use interrupt mode, the convergence time within 50 milliseconds.
- ☞ Generally, the port is configured as poll mode, interrupt mode is only applied to better performance environment, but the security of poll mode is better than interrupt mode, port-scan-mode {interrupt | poll} command can be consulted.
- ☞ In normal configuration, it still forms ring broadcast storm or ring block, please open debug function of primary node MRPP, and used show MRPP statistics command to observe states of primary node and transfer node and statistics information is normal or not, and then sends results to our Technology Service Center.

7.5 ULPP

7.5.1 Introduction to ULPP

Each ULPP group has two uplink ports, they are master port and slave port. The port may be a physical port or a port channel. The member ports of ULPP group have three states: Forwarding, Standby, Down. Normally, only one port at the forwarding state, the other port is blocked at the Standby state. When the master port has the link problem, the master port becomes down state, and the slave port is siwthed to forwarding state.

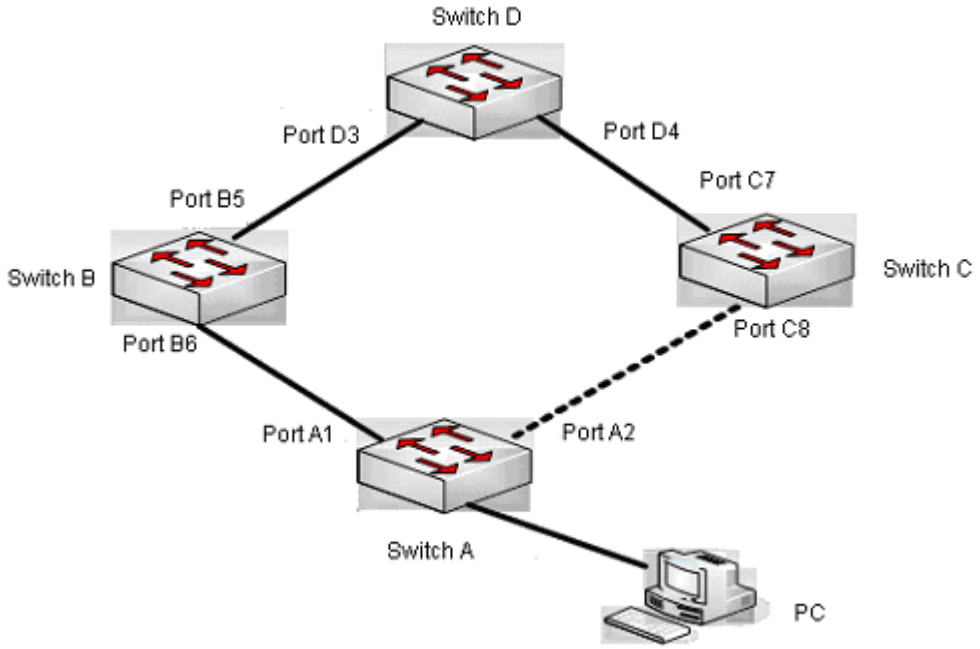


Figure 7-125 the using scene of ULPP

The above figure uses the double-uplink network, this is the typical application scene of ULPP. SwitchA goes up to SwitchD through SwitchB and SwitchC, port A1 and port A2 are the uplink ports. SwitchA configures ULPP, thereinto port A1 is set as the master port, port A2 is set as the slave port. When port A1 at forwarding state has the problem, switch the uplink at once, port A2 turns into forwarding state. After this, when recovering the master port, if the preemption mode is not configured, port A2 keeps the Forwarding state, port A1 turns into the Standby state.

After the preemption mode is enabled, so as to the master port preempts the slave port when it recovered from the problem. For avoiding the frequent uplink switch caused by the abnormity problem, the preemption delay mechanism is imported, and it needs to wait for some times before the master port preempt the slave port. For keeping the continuance of the flows, the master port does not process to preempt by default, but turns into the Standby state.

When configuring ULPP, it needs to specify the VLAN which is protected by this ULPP group through the method of MSTP instances, and ULPP does not provide the protection to other VLANs.

When the uplink switch is happennig, the primary forwarding entries of the device will not be applied to new topology in the network. In the figure, SwitchA configures ULPP, the portA1 as the master port at forwarding state, here the MAC address of PC is learned by Switch D from portD3. After this, portA1 has the problem, the traffic is switched to portA2 to be forwarded. If there is the data sent to PC by SwitchD, still the data will be forwarded from portD3, and will be losed. Therefore, when switching the uplink, the device of configuring ULPP needs to send the flush packets through the port which is switched to Forwarding state, and update MAC address tables and ARP tables of other devices in the network. ULPP respectively uses two kinds of flush packets to update the entries: the updated packets of MAC address and the deleted packets of ARP.

For making use of the bandwidth resource enough, ULPP can implement VLAN load balance through the configuration. As the picture illustrated, SwitchA configures two ULPP groups: portA1 is the master port and portA2 is the slave port in group1, portA2 is the master port and portA1 is the slave port in group2, the VLANs are protected by group1 and group2, they are 1-100 and 101-200. Here both portA1 and portA2 at the forwarding state, the master port and the slave port mutually backup, and respectively forward the packets of the different VLAN ranges. When portA1 has the problem, the traffic of VLAN 1-200 are forwarded by portA2. After this, when portA1 is recovering the normal state, portA2 forwards the data of VLAN 101-200 sequentially, but the data of VLAN 1-100 is switched to portA1 to forward.

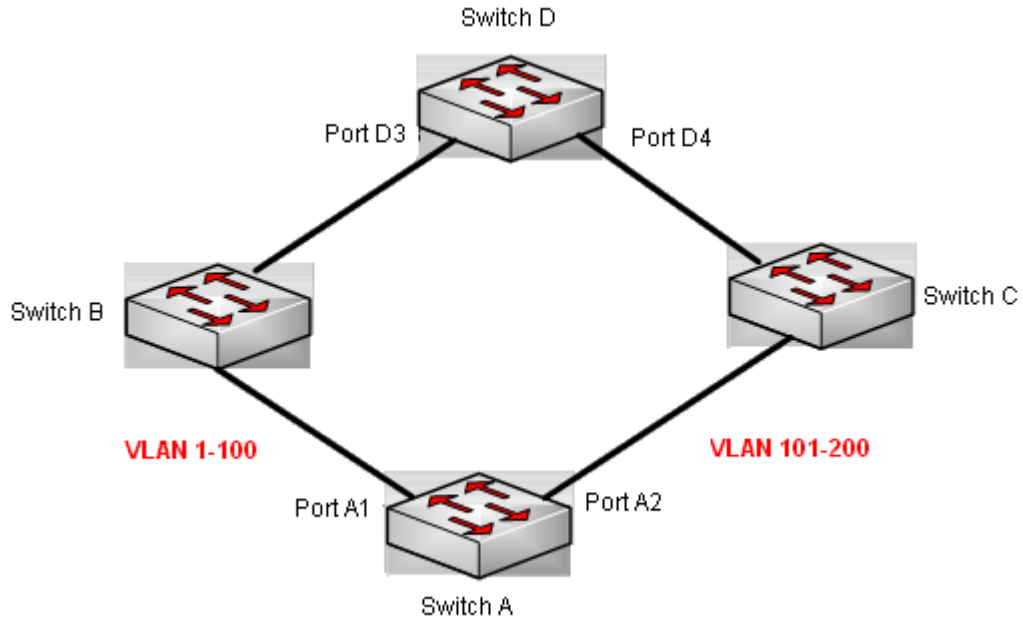


Figure 7-136 VLAN load balance

7.5.2 ULPP Configuration Task List

1. Create ULPP group globally
2. Configure ULPP group
3. Show and debug the relating information of ULPP

1. Create ULPP group globally

Command	Expalnation
Global mode	
ulpp group ***<integer>*** **no ulpp group** ***<integer>***	Configure and delete ULPP group globally.

2. Configure ULPP group

Command	Explanation
ULPP group configuration mode	

preemption mode **no preemption mode**	Configure the preemption mode of ULPP group. The no operation deletes the preemption mode.
preemption delay <*integer*> **no preemption delay**	Configure the preemption delay, the no operation restores the default value 30s.
control vlan <*integer*> **no control vlan**	Configure the sending control VLAN, no operation restores the default value 1.
protect vlan-reference-instance <*instance-list*> **no protect vlan-reference-instance <*instance-list*>**	Configure the protection VLANs, the no operation deletes the protection VLANs.
flush enable mac **flush disable mac**	Enable or disable sending the flush packets which update MAC address.
flush enable arp **flush disable arp**	Enable or disable sending the flush packets which delete ARP.
flush enable mac-vlan **flush disable mac-vlan**	Enable or disable sending the flush packets of deleting the dynamic unicast mac according to vlan.
description <*string*> **no description**	Configure or delete ULPP group description.
Port mode	
ulpp control vlan <*vlan-list*> **no ulpp control vlan <*vlan-list*>**	Configure the receiving control VLANs, no operation restores the default value 1.
ulpp flush enable mac **ulpp flush disable mac**	Enable or disable receiving the flush packets which update the MAC address.
ulpp flush enable arp **ulpp flush disable arp**	Enable or disable receiving the flush packets which delete ARP.
ulpp flush enable mac-vlan **ulpp flush disable mac-vlan**	Enable or disable receiving the flush packets of mac-vlan type.
ulpp group <*integer*> master **no ulpp group <*integer*> master**	Configure or delete the master port of ULPP group.
ulpp group <*integer*> slave **no ulpp group <*integer*> slave**	Configure or delete the slave port of ULPP group.

3. Show and debug the relating information of ULPP

Command	Explanation
Admin mode	
show ulpp group [group-id]	Show the configuration information of the configured ULPP group.

show ulpp flush counter interface {ethernet <IFNAME> | <IFNAME>}	Show the statistic information of the flush packets.
show ulpp flush-receive-port	Show flush type and control VLAN received by the port.
clear ulpp flush counter interface *<name>*	Clear the statistic information of the flush packets.
debug ulpp flush {send | receive} interface *<name>* **no debug ulpp flush {send | receive} interface *<name>***	Show the information of the receiving and sending flush packets, the no operation disables the shown information.
debug ulpp flush content interface *<name>* **no debug ulpp flush content interface *<name>***	Show the contents of the received flush packets, the no operation disables the showing.
debug ulpp error **no debug ulpp error**	Show the error information of ULPP, the no operation disables the showing.
debug ulpp event **no debug ulpp event**	Show the event information of ULPP, the no operation disables the showing.

7.5.3 ULPP Typical Examples

7.5.3.1 ULPP Typical Example1

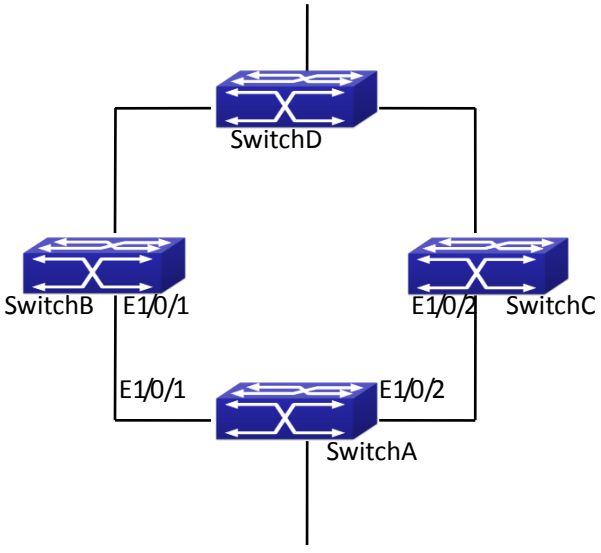


Figure 7-147 ULPP typical example1

The above topology is the typical application environment of ULPP protocol.

SwitchA has two uplinks, they are SwitchB and SwitchC. When any protocols are not enabled, this topology forms a ring. For avoiding the loopback, SwitchA can configure ULPP protocol, the master port and the slave port of ULPP group. When both master port and slave port are up, the slave port will be set as standby state and will not forward the data packets. When the master

port is down, the slave port will be set as forwarding state and switch to the uplink. SwitchB and SwitchC can enable the command that receives the flush packets, it is used to associate with ULPP protocol running of SwitchA to switch the uplink immediately and reduce the switch delay.

When configuring ULPP protocol of SwitchA, first, create a ULPP group and configure the protection VLAN of this group as vlan10, then configure interface Ethernet 1/0/1 as the master port, interface Ethernet 1/0/2 as the slave port, the control VLAN as 10. SwitchB and SwitchC configure the flush packets that receive ULPP.

SwitchA configuration task list:

```
Switch(Config)#vlan 10
Switch(Config-vlan10)#switchport interface ethernet 1/0/1; 1/0/2
Switch(Config-vlan10)#exit
Switch(Config)#spanning-tree mst configuration
Switch(Config-Mstp-Region)#instance 1 vlan 10
Switch(Config-Mstp-Region)#exit
Switch(Config)#ulpp group 1
Switch(ulpp-group-1)#protect vlan-reference-instance 1
Switch(ulpp-group-1)#control vlan 10
Switch(ulpp-group-1)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)# ulpp group 1 master
Switch(config-If-Ethernet1/0/1)#exit
Switch(Config)#interface Ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)# ulpp group 1 slave
Switch(config-If-Ethernet1/0/2)#exit
```

SwitchB configuration task list:

```
Switch(Config)#vlan 10
Switch(Config-vlan10)#switchport interface ethernet 1/0/1
Switch(Config-vlan10)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)# ulpp flush enable mac
Switch(config-If-Ethernet1/0/1)# ulpp flush enable arp
Switch(config-If-Ethernet1/0/1)# ulpp control vlan 10
```

SwitchC configuration task list:

```
Switch(Config)#vlan 10
Switch(Config-vlan10)#switchport interface ethernet 1/0/2
Switch(Config-vlan10)#exit
Switch(Config)#interface ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)# ulpp flush enable mac
Switch(config-If-Ethernet1/0/2)# ulpp flush enable arp
Switch(config-If-Ethernet1/0/2)# ulpp control vlan 10
```

7.5.3.2 ULPP Typical Example2

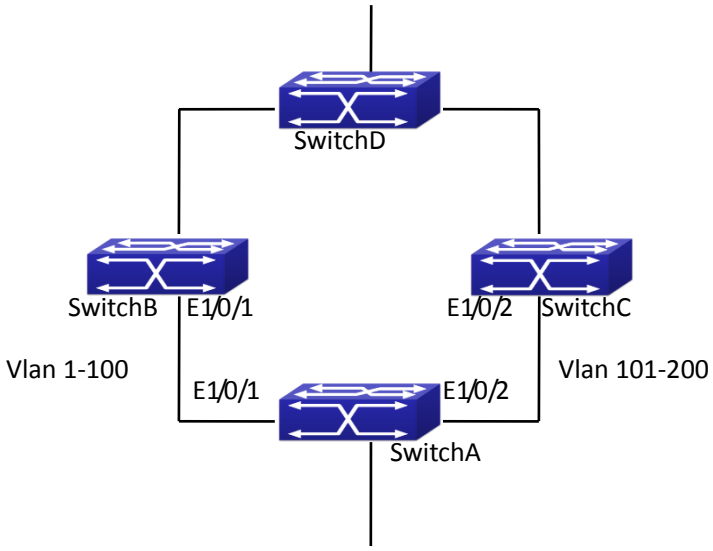


Figure 7-158 ULPP typical example2

ULPP can implement the VLAN-based load balance. As the picture illustrated, SwitchA configures two ULPP groups: port E1/0/1 is the master port and port 1/0/2 is the slave port in group1, port 1/0/2 is the master port and port 1/0/1 is the slave port in group2. The VLANs protected by group1 are 1-100 and by group2 are 101-200. Here both port E1/0/1 and port E1/0/2 at the forwarding state, the master port and the slave port mutually backup, respectively forward the packets of different VLAN ranges. When port E1/0/1 has the problem, the traffic of VLAN 1-200 are forwarded by port E1/0/2. When port E1/0/1 is recovering the normal state, still port E1/0/2 forwards the data of VLAN 101-200, the data of VLAN 1-100 are switched to port E1/0/1 to forward.

SwitchA configuration task list:

```
Switch(Config)#spanning-tree mst configuration
Switch(Config-Mstp-Region)#instance 1 vlan 1-100
Switch(Config-Mstp-Region)#instance 2 vlan 101-200
Switch(Config-Mstp-Region)#exit
Switch(Config)#ulpp group 1
Switch(ulpp-group-1)#protect vlan-reference-instance 1
Switch(ulpp-group-1)#preemption mode
Switch(ulpp-group-1)#exit
Switch(Config)#ulpp group 2
Switch(ulpp-group-2)#protect vlan-reference-instance 2
Switch(ulpp-group-1)#preemption mode
Switch(ulpp-group-2)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)#switchport mode trunk
Switch(config-If-Ethernet1/0/1)#ulpp group 1 master
Switch(config-If-Ethernet1/0/1)#ulpp group 2 slave
Switch(config-If-Ethernet1/0/1)#exit
```

Switch(Config)#interface Ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)#switchport mode trunk
Switch(config-If-Ethernet1/0/2)# ulpp group 1 slave
Switch(config-If-Ethernet1/0/2)# ulpp group 2 master
Switch(config-If-Ethernet1/0/2)#exit

SwitchB configuration task list:
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)#switchport mode trunk
Switch(config-If-Ethernet1/0/1)# ulpp flush enable mac
Switch(config-If-Ethernet1/0/1)# ulpp flush enable arp

SwitchC configuration task list:
Switch(Config)#interface ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)# switchport mode trunk
Switch(config-If-Ethernet1/0/2)# ulpp flush enable mac
Switch(config-If-Ethernet1/0/2)# ulpp flush enable arp

7.5.4 ULPP Troubleshooting

- At present, configuration of more than 2 multi-uplinks is allowed, but it may cause loopback, so is not recommended.
- With the normal configuration, if the broadcast storm happen or the communication along the ring is broken, please enable the debug of ULPP, copy the debug information of 3 minutes and the configuration information, send them to our technical service center.

7.6 ULSM

7.6.1 Introduction to ULSM

ULSM (Uplink State Monitor) is used to process the port state synchronization. Each ULSM group is made up of the uplink port and the downlink port, both the uplink port and the downlink port may be multiple. The port may be a physical port or a port channel, but it can not be a member port of a port channel, and each port only belongs to one ULSM group.

The uplink port is the monitored port of ULSM group. When all uplink ports are down or there is no uplink port in ULSM group, ULSM group state is down. ULSM group state is up as long as one uplink port is up.

The downlink port is the controlled port, its state changes along with Up/Down of ULSM group and is always the same with ULSM group state.

ULSM associates with ULPP to enable the downstream device to apperceive the link problem of the upstream device and process correctly. As the picture illustrated, SwitchA configures ULPP, here the traffic is forwarded by port A1. If the link between SwitchB and Switch D has the problem, SwitchA can not apperceive the problem of the upstream link and sequentially forward the traffic from port A1, cause traffic losing.

Configuring ULSM on SwitchB can solve the above problems. The steps are: set port B5 as the uplink port of ULSM group, port B6 as the downlink port. When the link between SwitchB and SwitchD has the problem, both the downlink port B6 and the state of ULSM group are down. It causes Switch A on which ULPP is configured to process uplink switchover and avoid the data dropped.

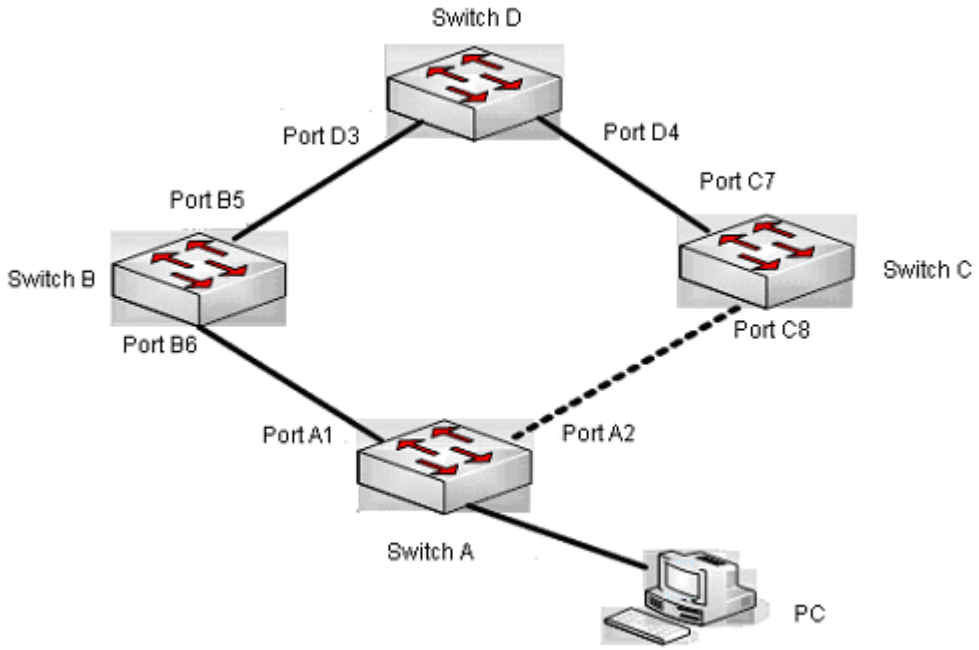


Figure 7-169 ULSM using scene

7.6.2 ULSM Configuration Task List

1. Create ULSM group globally
2. Configure ULSM group
3. Show and debug the relating information of ULSM

1. Create ULSM group globally

Command	explanation
Global mode	
ulsm group *<group-id>* **no ulsm group *<group-id>***	Configure and delete ULSM group globally.

2. Configure ULSM group

Command	explanation
Port mode	
ulsm group *<group-id>* {uplink | downlink} **no ulsm group *<group-id>* {uplink | downlink}**	Configure the uplink/downlink port of ULSM group, the no command deletes the uplink/downlink port.

3. Show and debug the relating information of ULSM

Command	Explanation
Admin mode	
show ulsm group [group-id]	Show the configuration information of ULSM group.
debug ulsm event **no debug ulsm event**	Show the event information of ULSM, the no operation disables the shown information.

7.6.3 ULSM Typical Example

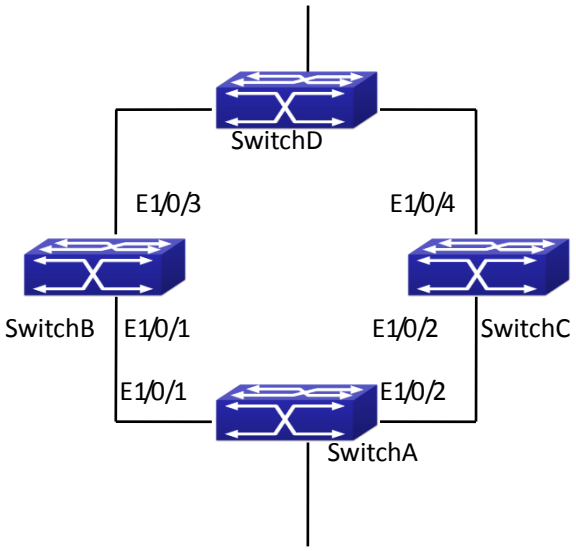


Figure 7-20 Fig 5-2 ULSM typical example

The above topology is the typical application environment which is used by ULSM and ULPP protocol.

ULSM is used to process the port state synchronization, its independent running is useless, so it usually associates with ULPP protocol to use. In the topology, SwitchA enables ULPP protocol, it is used to switch the uplink. SwitchB and SwitchC enable ULSM protocol to monitor whether the uplink is down. If it is down, then ULSM will execute the down operation for the downlink port to shutdown it, so ULPP protocol of Swtich A executes the relative operation of the uplink switchover.

SwitchA configuration task list:

Switch(Config)#spanning-tree mst configuration

Switch(Config-Mstp-Region)#instance 1 vlan 1

```
Switch(Config-Mstp-Region)#exit
Switch(Config)#ulpp group 1
Switch(ulpp-group-1)#protect vlan-reference-instance 1
Switch(ulpp-group-1)#exit
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)# ulpp group 1 master
Switch(config-If-Ethernet1/0/1)#exit
Switch(Config)#interface Ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)# ulpp group 1 slave
Switch(config-If-Ethernet1/0/2)#exit
```

SwitchB configuration task list:

```
Switch(Config)#ulsm group 1
Switch(Config)#interface ethernet 1/0/1
Switch(config-If-Ethernet1/0/1)#ulsm group 1 downlink
Switch(config-If-Ethernet1/0/1)#exit
Switch(Config)#interface ethernet 1/0/3
Switch(config-If-Ethernet1/0/3)#ulsm group 1 uplink
Switch(config-If-Ethernet1/0/3)#exit
```

SwitchC configuration task list:

```
Switch(Config)#ulsm group 1
Switch(Config)#interface ethernet 1/0/2
Switch(config-If-Ethernet1/0/2)#ulsm group 1 downlink
Switch(config-If-Ethernet1/0/2)#exit
Switch(Config)#interface ethernet 1/0/4
Switch(config-If-Ethernet1/0/4)#ulsm group 1 uplink
Switch(config-If-Ethernet1/0/4)#exit
```

7.6.4 ULSM Troubleshooting

☞ With the normal configuration, if the downlink port does not responds the down event of the uplink port, please enable the debug function of ULSM, copy the debug information of 3 minutes and the configuration information, and send them to our technical service center.

Chapter 8 Debugging and Diagnosis Configuration

8.1 Monitor and Debug

When the users configures the switch, they will need to verify whether the configurations are correct and the switch is operating as expected, and in network failure, the users will also need to diagnostic the problem. Switch provides various debug commands including ping, telnet, show and debug, etc. to help the users to check system configuration, operating status and locate problem causes.

8.1.1 Ping

Ping command is mainly used for sending ICMP query packet from the switches to remote devices, also for check the accessibility between the switch and the remote device. Refer to the Ping command chapter in the Command Manual for explanations of various parameters and options of the Ping command.

8.1.2 Ping6

Ping6 command is mainly used by the switch to send ICMPv6 query packet to the remote equipment, verifying the accessibility between the switch and the remote equipment. Options and explanations of the parameters of the Ping6 command please refer to Ping6 command chapter in the command manual.

8.1.3 Icmp-reply fast

The icmp-reply fast command enables the switch to quickly reply ICMP Echo Request packets to shorten the reply time.

8.1.4 Traceroute

Traceroute command is for testing the gateways through which the data packets travel from the source device to the destination device, so to check the network accessibility and locate the network failure.

Execution procedure of the Traceroute command consists of: first a data packet with TTL at 1 is sent to the destination address, if the first hop returns an ICMP error message to inform this packet can not be sent (due to TTL timeout), a data packet with TTL at 2 will be sent. Also the send hop may be a TTL timeout return, but the procedure will carries on till the data packet is

sent to its destination. These procedures is for recording every source address which returned ICMP TTL timeout message, so to describe a path the IP data packets traveled to reach the destination.

Traceroute Options and explanations of the parameters of the Traceroute command please refer to traceroute command chapter in the command manual.

8.1.5 Traceroute6

The Traceroute6 function is used on testing the gateways passed through by the data packets from the source equipment to the destination equipment, to verify the accessibility and locate the network failure. The principle of the Traceroute6 under IPv6 is the same as that under IPv4, which adopts the hop limit field of the ICMPv6 and IPv6 header. First, Traceroute6 sends an IPv6 datagram (including source address, destination address and packet sent time) whose HOPLIMIT is set to 1. When first route on the path receives this datagram, it minus the HOPLIMIT by 1 and the HOPLIMIT is now 0. So the router will discard this datagram and returns with a 「ICMPv6 time exceeded」 message (including the source address of the IPv6 packet, all content in the IPv6 packet and the IPv6 address of the router). Upon receiving this message, the Traceroute6 sends another datagram of which the HOPLIMIT is increased to 2 so to discover the second router. Plus 1 to the HOPLIMIT every time to discover another router, the Traceroute6 repeat this action till certain datagram reaches the destination.

Traceroute6 Options and explanations of the parameters of the Traceroute6 command please refer to traceroute6 command chapter in the command manual.

8.1.6 Show

show command is used to display information about the system, port and protocol operation. This part introduces the **show** command that displays system information, other **show** commands will be discussed in other chapters.

Command	Explanation
Admin Mode	
show debugging	Display the debugging state.
show flash	Display the files and the sizes saved in the flash.
show history	Display the recent user input history command.
show history all-users [detail]	Show the recent command history of all users. Use **clear history all-users** command to clear the command history of all users saved by the system, the max history number can be set by **history all-users max-length** command.
show memory usage	Show the memory usage.
show running-config	Display the switch parameter configuration validating at current operation state.
show running-config current-mode	Show the configuration under the current mode.

show startup-config	Display the switch parameter configuration written in the Flash Memory at current operation state, which is normally the configuration file applied in next time the switch starts up.
show switchport interface [ethernet <*IFNAME*>]	Display the VLAN port mode and the belonging VLAN number of the switch as well as the Trunk port information.
show tcp **show tcp ipv6**	Display the TCP connection status established currently on the switch.
show udp **show udp ipv6**	Display the UDP connection status established currently on the switch.
show telnet login	Display the information of the Telnet client which currently establishes a Telnet connection with the switch.
show tech-support	Display the operation information and the state of each task running on the switch. It is used by the technicians to diagnose whether the switch operates properly.
show version	Display the version of the switch.
show temperature	Show CPU temperature of the switch.
show fan	Show fan information of switch

8.1.7 Debug

All the protocols switch supports have their corresponding debug commands. The users can use the information from debug commands for troubleshooting. Debug commands for their corresponding protocols will be introduced in the later chapters.

8.2 Logging

All the protocols switch supports record the commands executed by user at the console, telnet or ssh terminal and send the log to info-center.

Command	Description
Global mode	
logging executed-commands {enable | disable}	Enable or disable the logging executed-commands
Admin mode	
show logging executed-commands state	Show the state of logging executed-commands

8.2.1 System Log Introduction

8.2.1.1 Log Output Channel

8.2.1.2 Format and Severity of the Log Information

8.2.2 System Log Configuration

8.2.3 System Log Configuration Example

8.3 Reload Switch after Specified Time

8.3.1 Introduce to Reload Switch after Specifid Time

Reload switch after specified time is to reboot the switch without shutdown its power after a specified period of time, usually when updating the switch version. The switch can be rebooted after a period of time instead of immediately after its version being updated successfully.

8.3.2 Reload Switch after Specifid Time Task List

1. Reload switch after specified time

Command	Explanation
Admin mode	
reload after {[<*HH:MM:SS*>] [days <*days*>]}	Reload the switch after a specified time period.
reload cancel	Cancel the specified time period to reload the switch.

8.4 Debugging and Diagnosis for Packets Received and Sent by CPU

8.4.1 Introduction to Debugging and Diagnosis for Packets Received and Sent by CPU

The following commands are used to debug and diagnose the packets received and sent by CPU, and are supposed to be used with the help of the technical support.

8.4.2 Debugging and Diagnosis for Packets Received and Sent by CPU Task List

Command	Explanation
Global Mode	
cpu-rx-ratelimit total *<packets>* **no cpu-rx-ratelimit total**	Set the total rate of the CPU receiving packets, the no command sets the total rate of the CPU receiving packets to default.
cpu-rx-ratelimit protocol *<protocol-type> <packets>* **no cpu-rx-ratelimit protocol [*<protocol- type>*]**	Set the max rate of the CPU receiving packets of the protocol type, the no command set the max rate to default.
clear cpu-rx-stat protocol [*<protocol-type>*]	Clear the statistics of the CPU received packets of the protocol type.
Admin Mode	
show cpu-rx protocol [*<protocol-type>*]	Show the information of the CPU received packets of the protocol type.
debug driver {receive|send} [interface {*<interface-name>* |all}] [protocol {*<protocol-type>* |discard |all}] [detail]	Turn on the showing of the CPU receiving or sending packet informations.
no debug driver {receive | send}	Turn off the showing of the CPU receiving or sending packet informations.

Command	Explanation
Admin Mode	
protocol filter {protocol-type}	Turn on/off the treatment of the named protocol packets, the

no Protocol filter {protocol-type}	named protocol contains: **{arp|bgp|dhcp|dhcpv6|hsrp|http|igmp|ip|ldp|mpls|ospf|pim|rip|snmp|telnet|vrrp}**

8.5 DCP

8.5.1 Introduction to DCP

The dynamic CPU protection is also named as dynamic CPU limit-rate. When the rate of the packet with the special source IP going up the CPU is detected exceeding the certain value, these packets will be limited the rate. The allowed maximum rate of the packet going up the CPU is named as limit-rate. The limit-rate can be the configured default value or the configured value by user.

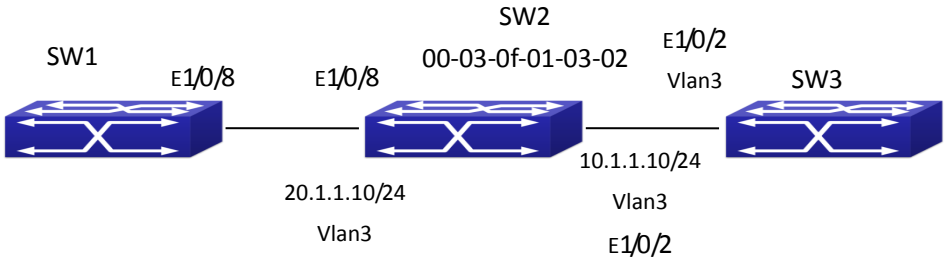
DCP is used to limit the rate for the network segment IP to prevent one IP occupying the bandwidth. DCP (Dynamic CPU Protection) means to control the rate of the packet going up the CPU through monitoring the other-ipuc packet going up the CPU for avoiding that the rate is too fast and causes overload. It can protect the CPU. When the flow is less than the half of the limit-rate in 5s, the rate limiting will be canceled. DCP is only for the other-ipuc packet, for the management packet or protocol packet, DCP is not adopted.

8.5.2 DCP Configuration

Command	Explanation
Global Configuration Mode	
dcp enable **dcp disable**	Enable/disable the dcp function.
dcp limit-rate <20-50> **no dcp limit-rate**	Configure the limit-rate value of dcp. The no command cancels it and recovers it to be the default value.
dcp no-limit-ip <ip_addr> **no dcp no-limit-ip <ip_addr>**	Configure the IP that the dcp does not limit its rate. The no command cancels it.
show dcp limit-rate	Show the limit-rate configured by user.

show cpu ip rate top10 [slot <1-9>|member <1-16>]	Show the first 10 IP with the maximum rate of going on cpu in 5s and show the limit-rate value.
show dcp limited ip [slot <1-9>|member <1-16>]	Show the node information of the ip which is limited the rate.
Admin Mode	
clear dcp speed limit rules {member <1-16>}	Clear the rate limiting rule that the DCP sent to the drive.
debug dcp packet **no debug dcp packet**	Show the process that the DCP deals with and monitor the packet going up the CPU, the no command cancels printing.
debug dcp event **no debug dcp event**	Show the process that the DCP deals with the events. The no command cancels printing.

8.5.3 DCP Configuration Examples



As shown in the above topology, send from E1/0/8 of SW1 to SW2, the destination mac is the one of SW2 which is 00-03-0f-01-03-02, the non-protocol and non-management packet with the destination ip of 10.1.1.X/24 (this ip address cannot be achieved) will be identified as other-ipuc packet. This packet will be sent to CPU. When a lot of packets like this are sent to the CPU, the CPU will be under the heavy load and it will cause that the normal business cannot be dealt with.

If enabled the DCP, when the rate of the packet with the special source IP going up the CPU is detected exceeding the certain value, these packets will be limited the rate. The CPU can be protected.

Configuration:

1. Enable DCP
2. Configure the limit-rate
3. Configure the IP address with no rate limiting
4. Show the configured limit-rate
5. Show the first 10 IP with maximum rates of going up the CPU in 5s and show the

limit-rate

6. Show the node information of the ip which is limited the rate
7. Clear the rate limiting rule that the DCP sent to the drive
8. Show the process that the DCP deals with and monitor the packet going up the CPU
9. Cancel the process that the DCP deals with and monitor the packet going up the CPU
10. Show the process that the DCP deals with the events
11. Cancel printing of the process that the DCP deals with the events

Configuration steps:

Switch(Config)# dcp enable

Switch(Config)# dcp limit-rate 50

Switch(Config)# dcp no-limit-ip 1.1.1.1

Switch(config)#show dcp limit-rate

Switch(config)#show cpu ip rate top10

Switch(config)#show dcp limited ip

Switch#clear dcp speed limit rules
Switch#debug dcp packet
Switch#no debug dcp packet
Switch#debug dcp event
Switch#no debug dcp event

8.5.4 DCP Configuration Troubleshooting

Please pay attention to the following points when using and configuring DCP:

- Under the default configuration, dcp is disabled. It can be effective only after configured the command of dcp enable.
- DCP is only for the other-ipuc packet, for the management packet or protocol packet, DCP is not adopted.
- DCP no-limited-ip can configure the maximum of 1024, if exceeds this value, it cannot be issued.
- Under the default configuration, the limit-rate for other-ipuc packet is 20.
- When the flow is less than the half of the limit-rate in 5s, the rate limiting will be canceled.
- Show the first 10 IP with maximum rates of going up the CPU in 5s and show the limit-rate through the command of show cpu ip rate top10.
- Show the node information of the ip which is limited the rate through the command of show dcp limited ip. Limited-IP is the ip which is limited the rate, Rate(pkts/s) is the current rate.
- Enable the on-off of debug dcp packet or debug dcp event to view the process that the DCP deals with the packet going up the CPU. The command of debug dcp packet can be used to view the detailed information of the packet including source IP, destination IP, source port,

destination port, protocol number, etc. The command of debug dcp event can be used to print the process that the DCP deals with the events.

8.6 COPP

8.6.1 Introduction to COPP

CPU is the brain of the device and it can deal with the information of all control planes. So the CPU should be protected through the appropriate measures. The CPU rate limiting keeps the previous protocol rate limiting function, and CPU rate limiting is changed to hardware rate limiting. The new function of COPP (control plane policing) can protect the control and management panel for ensuring the stability of the routing function and the normal transmission of the packets. Configure the ACL first, multiple ACL rules are supported, classify the specific packets which should be limited rate or filtered through the ACL rules. And then configure the ACL rule matching to the COPP policy map to filter the specific packets or limit the rate. The COPP supports single bucket mode, dual bucket mode and multiple configurations.

8.6.2 COPP Configuration

1. Configuire the ACL rule, the digital standard IP access-list is as the example
2. Create the class-map
3. Create the copp-policy-map
4. Create the policy
5. Apply to the port

1. Configuire the ACL rule

Command	Explanation
Global Configuration Mode	
access-list *<num>* {deny | permit} {{*<sIpAddr>* *<sMask>*} | any-source | {host-source *<sIpAddr>*}} **no access-list *<num>***	Create a digital standard IP access list. If this list has existed, add a rule entry. The no command deletes the access list.

2. Create the class-map

Command	Explanation
Global Configuration Mode	
class-map *<class-map-name>* **no class-map *<class-map-name>***	Create a class-map and enter the class-map mode. The no command deletes the appointed class-map.
match {access-group <acl-index-or-name> | ip	Configure the matching standard of the

dscp <dscp-list>| ip precedence <ip-precedence-list>| ipv6 access-group <acl-index-or-name> | ipv6 dscp <dscp-list> | ipv6 flowlabel <flowlabel-list> | vlan <vlan-list> / cos *<cos-list>*| vlan range <vlan-list>} **no match {access-group | ip dscp | ip precedence / ipv6 access-group | ipv6 dscp | ipv6 flowlabel | vlan |cos| vlan range}**	class-map and classify the data according to ACL. The no command deletes it.

3. Create the copp-policy-map

Command	Explanation
Global Configuration Mode	
copp-policy-map <*policy-map name*> **no policy-map <*policy-map-name*>**	Create a copp-policy-map and enter the copp-policy-map mode. The no command deletes the appointed copp-policy-map.
class <*class-map-name*> **no class <*class-map-name*>**	After create a copp-policy-map, it can be corresponding to a class, and different policies can be adopted for different data flow after enter the policy class map configuration mode. The no command deletes the appointed policy class map.
pps mode: **policy packets-per-second <pps> normal-burst-packets <pps> { conform-action | exceed-action } <ACTION>** **no policy** **bps mode:** **1. Single Bucket Mode:** **Policy <bits_per_second> <normal_burst_bytes> ({action{{policied-cos-to-cos-transmit{policied-cos-to-dscp-transmit|violate-action}|policied-cos-to-dscp-transmit{policied-cos-to-cos-transmit|violate-action }| policied-dscp-exp-to-cos-transmit{policied-dscp-exp-to-dscp-transmit|violate-action}|policied-dscp-exp-to-dscp-transmit{policied-dscp-exp-to-cos-transmit| violate-action }}violate-action {drop| transmit}} | exceed-action ACTION })** **2. Dual Bucket Mode:** **policy <bits_per_second>**	pps mode: It supports the policy command of single bucket two colors and the limit-rate is pps mode, divide the packets into different colors according to the configuration, and set the corresponding action for different color packets. The no operation will delete the mode configuration. bps mode: It supports the non-aggregation policy command of three colors, analyze the working mode of the token bucket, whether it is single rate singe bucket, singe rate dual bucket or dual rate dual bucket, and set the corresponding action for different color packets. The no operation will delete the mode configuration.

**<normal_burst_bytes> [pir <peak_rate_bps>]	<maximum_burst_bytes> [{action{{policied-cos-to-cos-transmit{policied-cos-to-dscp-transmit|violate-action}|policied-cos-to-dscp-transmit{policied-cos-to-cos-transmit|violate-action }| policied-dscp-exp-to-cos-transmit{policied-dscp-exp-to-dscp-transmit|violate-action}|policied-dscp-exp-to-dscp-transmit{policied-dscp-exp-to-cos-transmit| violate-action }}	exceed-action	violate-action ACTION }]** **ACTION definition:** **drop	transmit	set-internal-priority <intp_value>	policied-intp-transmit** **no policy**	
Policy Class Map Configuration Mode	
drop **no drop** **transmit** **no transmit**	Choose dropping or transmitting for the classified flow. The no command cancels it.

4. Apply to the port

Command	Explanation
Port Configuration Mode	
service-policy output *<policy-map name>* **no service-policy output *<policy-map-name>***	Apply a policy map to the egress of the port. The no command deletes the policy map.

8.6.3 COPP Configuration Examples

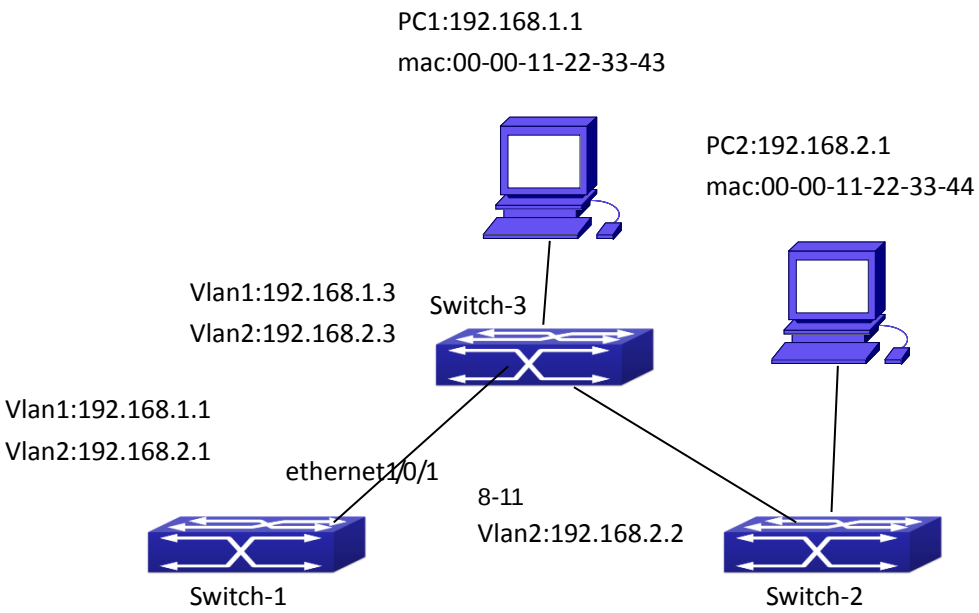


Figure 8-1 COPP function

Example 1: Configure the COPP policy map on ethernet1/0/1 of Switch-1, limit the packets in 192.168.1.0 as 10pps, and configure the burst threshold as 20pps. Drop all the packets whose bandwidth exceeds the value.

Configuration steps:

```
Switch#config
Switch(config)#access-list 1 permit 192.168.1.0 0.0.0.255
Switch(config)#class-map c1
Switch(config-classmap-c1)#match access-group 1
Switch(config-classmap)# exit
Switch(config)#copp-policy-map p1
Switch(config-copp-policymap-p1)#class c1
Switch(config-copp-policymap-p1-class-c1)#policy packets-per-second 10 normal-burst-packets 20 exceed-action drop
Switch(config-copp-policymap-p1-class-c1)#exit
Switch(config-copp-policymap-p1)# #exit
Switch(config)#interface ethernet 1/0/1
Switch(config-if-ethernet1/0/1)# service-policy output p1
```

Example 2: Configure the COPP policy map on ethernet1/0/1 of Switch-1, limit the packets of the source mac address of 00-00-11-22-33-44 as 10pps, and configure the burst threshold as 20pps. Configure the priority as 1 and forward the packets whose bandwidth exceeds the value.

Configuration steps:

```
Switch#config
Switch(config)#access-list 1100 permit host-source-mac 00-00-11-22-33-44 any-destination-mac
Switch(config)#class-map c1
Switch(config-classmap-c1)#match access-group 1100
Switch(config)#copp-policy-map p1
Switch(config-copp-policymap-p1)#class c1
Switch(config-copp-policymap-p1-class-c1)# policy packets-per-second 10 normal-burst-packets 20 exceed-action set-internal-priority 1 transmit
Switch(config-copp-policymap-p1-class-c1)#exit
Switch(config-copp-policymap-p1)# #exit
Switch(config)#interface ethernet 1/0/1
Switch(config-if-ethernet1/0/1)# service-policy output p1
```

8.6.4 COPP Configuration Troubleshooting

☞ If the packets which belong to this network segment went on the cpu correctly. This function can limit the rate of the packets going on the cpu, user can check if the packets went on the

cpu through the command of debug driver receive.

- ☞ If the ACL is matched correctly. The ACL rule must be permit.
- ☞ If the COPP policy map is configured correctly. COPP must be on the egress direction on the port, it does not support the ingress direction.
- ☞ User can view the packets statistics and check if the rate limiting is effective through the command of show cpu-rx protocol all.

8.7 Info-center

8.7.1 Overview

8.7.1.1 Basic Concepts

The information center is a set of mechanisms that outputs system information (hereafter referred to as the information source).

Users need to configure for the output direction, and enable the corresponding output direction, and set matching information conditions to output the desired log information, the configuration process is simple and flexible.

8.7.1.2 Information sources

There are eight levels of severity for information sources.

Table 1-1

Grade num	Grade name	Description
1	emergencies	Fatal error
2	alerts	Errors to be corrected immediately
3	critical	Critical Error
4	errors	Concerned but not critical errors
5	warnings	Warning, there may be some error
6	notifications	Note the information
7	informational	General information
8	debugging	Debugging information

8.7.1.3 Output direction

Currently, 16 output directions are supported, as shown in Table 1-2:

Table 1-2

Output direction	Description
console	Control panel
monitor	monitor terminal
loghost <1-8>	Log host
trapbuffer	Alarm buffer
logbuffer	Log buffer
logfile <1-4>	Log file

The Settings for the output directions are independent of each other, but the information center needs to be opened before the settings take effect.

8.7.1.4 Source information format

The format of the source information is as follows：

<priority>timestamp sysname module/level/digest:content

The angle brackets (<>), spaces, slashes (/), and colon (:) are required in the above format when configuring the log prefix; where angle brackets are valid only when the output direction is the log host.

An example of the log format to be output to the log host is shown below:

<188>Sep 28 15:33:46:235 2005 MyDevice SHELL/5/LOGIN: Console login from con0

Another example of the log format without prefix is shown below:

Console login from con0

Each field is described in detail below:

1) Priority

The priority is calculated as follows: facility*8+severity-1。

Facility can be local0 ~ local7, default is local7. The value is 16-23.

Severity ranges from 1 to 8. For more information about severity, refer to Table 1-1.

There is no character between the priority and the timestamp. The priority is valid only when the information is sent to the log host.

The origin of facililty is worth mentioning. Facility is a unique attribute when information is output to the log host, and has no meaning for other output directions. Its value local0 ~ local7 comes from the log type which the early syslog protocol defined for the UNIX log host. All types are as follows:

Table 1-3

facility	code
kernel messages	0
user-level messages	1
mail system	2
system daemons	3

security/authorization messages (note 1)	4
messages generated internally by syslogd	5
line printer subsystem	6
network news subsystem	7
UUCP subsystem	8
clock daemon (note 2)	9
security/authorization messages (note 1)	10
FTP daemon	11
NTP subsystem	12
log audit (note 1)	13
log alert (note 1)	14
clock daemon (note 2)	15
local use 0 (local0)	16
local use 1 (local1)	17
local use 2 (local2)	18
local use 3 (local3)	19
local use 4 (local4)	20
local use 5 (local5)	21
local use 6 (local6)	22
local use 7 (local7)	23

As you can see, local0 ~ local7 is reserved for local users, the value is 16 to 23.

2) Timestamp

The timestamp records the time when the system information is generated, making it easy for the user to view and locate system events. The timestamp and host name are separated by a space.

Date mode

mmm + space + dd + space + hh + colon +mm + colon + ss + colon + sss+ space+ yyyy

mmm = Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec

Indicates month

dd = 1~31, Indicates the date, if the value is 1 to 9, 0 should be added in front of the number. For example, 1 to be converted to "01".

hh = 0 ~ 23, Indicates hours

mm = 0 ~ 59, Indicates minutes

ss = 0 ~ 59 Indicates seconds

sss = 0 ~ 999 Indicates milliseconds

yyyy = 1970 ~ 9999 Indicates year

3) Sysname

The sysname is the system name of the machine.

The sysname and module name are separated by a space.

4) module

This field indicates the name of the module that generated the message.

The module and the level are separated by a slash (/).

5) Severity

System information is divided into three categories: log, debug and alarm. According to the urgency, each type of information is divided into eight levels. The smaller information level value, the higher urgency level.

The rule that the system uses to filter information is to prohibit the output of information whose information level is greater than threshold.

When the value of the information level is 1, the system will only output emergencies information;

When the value of the information level is 8, the system outputs all levels of information;

If the configuration filtering method is an exact match, the system only outputs information at the design level.

The level is separated from the digest by a slash (/).

6) digest

A message digest is a string of up to 32 characters that indicates the content of the message.

The digest is separated from the content by a colon (:).

7) content

8.7.2 The content indicates the specific contents of the system information.Configure Sequence

1. Set the global enable of the information center
2. Set the output direction of the information center to enable
3. Set the log information prefix switch
4. Set the log information matching conditions
5. Configure the IP and facility of the log host
6. Configure the number of log files and access path
7. Configure to record user operation commands
8. Delete logs recorded by logbuffer or trapbuffer
9. View the area that supports storing files
10. Configure onekey collection

1. Set the global enable of the information center

Command	Description
Global configuration mode	
info-center enable	Set the global enable of the information center
no info-center enable	Set the global disable of the information center

2. Set the output direction of the information center to enable

Command	Description

Global configuration mode	
info-center (console | logbuffer | monitor | trapbuffer) output-enable info-center (logfile <1-4> | loghost <1-8>) output-enable (member <1-4> |)	Set the output direction of the information center to enable
no info-center (console | logbuffer | monitor | trapbuffer) output-enable no info-center (logfile <1-4> | loghost <1-8>) output-enable (member <1-16> |)	Set the output direction of the information center to disable

3. Set the log information prefix switch

Command	Description
Global configuration mode	
info-center (console | logbuffer | monitor | trapbuffer) prefix (on|off) info-center (logfile <1-4> | loghost <1-8>) prefix (on|off) (member <1-16> |)	Set the log information prefix switch

4. Set the log information matching conditions

Command	Description
Global configuration mode	
info-center (console | logbuffer | monitor | trapbuffer) match level (emergencies | alerts | critical | errors | warnings | notifications | informational | debugging) (exact |) (keyword WORD |) info-center (logfile <1-4> | loghost <1-8>) match level (emergencies | alerts | critical | errors | warnings | notifications | informational | debugging) (exact |) (keyword WORD |) (member <1-16> |)	Set the log information matching conditions
no info-center (console | logbuffer | monitor | trapbuffer) match no info-center (logfile <1-4> | loghost <1-8>) match (member <1-16> |)	Delete the log information matching conditions

5. Configure the IP and facility of the log host

Command	Description
Global configuration mode	
info-center loghost <1-8> config (A.B.C.D | X:X::X:X) facility (local0|local1|local2|local3|local4|local5|local6|local7) (member <1-16> |)	Configure the IP and facility of the log host
no info-center loghost <1-8> config (member <1-16> |)	Delete the IP and facility of the log host

6. Configure the number of log files and access path

Command	Description
Global configuration mode	
info-center logfile <1-4> config count <1-40960> (flash|usb|nandflash) WORD (member <1-16> |)	Configure the number of log files and access path
no info-center logfile <1-4> config (member <1-16> |)	Delete the number of log files and access path

7. Configure to record user operation commands

Command	Description
Global configuration mode	
info-center (logbuffer | logfile <1-4> | loghost <1-8>) record-cmd	Configure to record user operation commands
no info-center (logbuffer | logfile <1-4> | loghost <1-8>) record-cmd	Delete to record user operation commands

8. Delete logs recorded by logbuffer or trapbuffer

Command	Description
Global configuration mode	
info-center clear trapbuffer	Delete logs recorded by trapbuffer
info-center clear logbuffer (member <1-16> |)	Delete logs recorded by logbuffer

9. View the area that supports storing files

Command	Description
Global configuration mode	
info-center list all disk	View the area that supports storing files

10. Configure onekey collection

Command	Description
Global configuration mode	
info-center save all (((flash|usb|nandflash) WORD)|)	Configure onekey collection

8.7.3 Instructions

8.7.3.1 console output log

Example 1, configuring the console is allowed to output the logs with prefix, and matching

level of debugging or above.

The configuration steps are as follows:
Switch(config)#info-center console output-enable
Switch(config)#info-center console prefix on
Switch(config)#info-center console match level debugging

8.7.3.2 monitor output log

Example 2, configuring the monitor is allowed to output the logs without prefix, and matching level of warnings only.

The configuration steps are as follows:
Switch(config)#info-center monitor output-enable
Switch(config)#info-center monitor prefix off
Switch(config)#info-center monitor match level warnings exact

8.7.3.3 logbuffer output log

Example 3, configuring the logbuffer is allowed to output the logs with prefix, and matching level of debugging or above, also matching regular expression DEVSM.

The configuration steps are as follows:
Switch(config)#info-center logbuffer output-enable
Switch(config)#info-center logbuffer prefix on
Switch(config)#info-center logbuffer match level debugging keyword DEVSM

8.7.3.4 trapbuffer output log

Example 4, configuring the trapbuffer is allowed to output the logs with prefix, and matching level of warnings only, also matching regular expression DEVSM.

The configuration steps are as follows:
Switch(config)#info-center trapbuffer output-enable
Switch(config)#info-center trapbuffer prefix on
Switch(config)#info-center trapbuffer match level warnings exact keyword DEVSM

8.7.3.5 loghost output log

Example 5, configuring the loghost 1 allowed to output the logs with prefix, and matching level of debugging or above, also the log host's IP is 192.168.1.11, facility is local0.

The configuration steps in vsf mode are as follows:
Switch(config)#info-center loghost 1 output-enable member 1

Switch(config)#info-center loghost 1 prefix on member 1
Switch(config)#info-center loghost 1 match level debugging member 1
Switch(config)#info-center loghost 1 config 192.168.1.11 facility local0 member 1

The configuration steps in stand-alone mode are as follows:
Switch(config)#info-center loghost 1 output-enable
Switch(config)#info-center loghost 1 prefix on
Switch(config)#info-center loghost 1 match level debugging
Switch(config)#info-center loghost 1 config 192.168.1.11 facility local0

8.7.3.6 logfile output log

Example 6, configuring the logfile 1 allowed to output the logs with prefix, and matching level of debugging or above, also up to 2000 logs are allowed. The storage disk is flash, the file name is file1.log.

The configuration steps in vsf mode are as follows:
Switch(config)#info-center logfile 1 output-enable member 1
Switch(config)#info-center logfile 1 prefix on member 1
Switch(config)#info-center logfile 1 match level debugging member 1
Switch(config)#info-center logfile 1 config count 2000 flash file1.log member 1

The configuration steps in stand-alone mode are as follows:
Switch(config)#info-center logfile 1 output-enable
Switch(config)#info-center logfile 1 prefix on
Switch(config)#info-center logfile 1 match level debugging
Switch(config)#info-center logfile 1 config count 2000 flash file1.log

8.7.3.7 record user operation commands

Example 7, configuring the logbuffer is allowed to record user operation commands.

The configuration steps are as follows:
Switch(config)#info-center logbuffer output-enable
Switch(config)#info-center logbuffer record-cmd

8.7.3.8 Delete logs recorded by logbuffer or trapbuffer

Example 8, delete logs recorded by logbuffer or trapbuffer.

The configuration steps are as follows:
Switch(config)#info-center clear logbuffer

Switch(config)#info-center clear trapbuffer

8.7.3.9 View the area that supports storing files

Example 9, view the area that supports storing files.

The configuration steps are as follows:
Switch(config)#info-center list all disk
flash:
usb:
nandflash:

8.7.3.10 Configure onekey collection

Example 10, configure onekey collection.

The configuration steps are as follows:
Switch(config)#info-center list all disk
*****Now saving Master card(member 1)*****
Now saving infocenter all configuration, please wait..
Now saving infocenter logbuffer content, please wait..
Now saving infocenter trapbuffer content, please wait..
*****Master card(member 1) saving finished!*****

*****Now saving Slave card(member 2)*****
Now saving infocenter logbuffer content, please wait..
*****Slave card(member 2) saving finished!*****

8.7.4 Erratum

- Configuring the matching level of log information is valid for subsequent logs, and the previously saved log information remains unchanged.
- Disk capacity is limited, and when the maximum number of logs in disk is reached, the new log information replaces the old log.
- The information center is globally enabled by default, and all current configurations of the information center can be displayed using the show info-center config command.

8.8 Mirror

8.8.1 Introduction to Mirror

Mirror functions include port mirror function, CPU mirror function, flow mirror function.

Port mirror refers to the duplication of data frames sent/received on a port to another port. The duplicated port is referred to as mirror source port and the duplicating port is referred to as mirror destination port. A protocol analyzer (such as Sniffer) or a RMON monitor will be connected at mirror destination port to monitor and manage the network, and diagnose the problems in the network.

CPU mirror function means that the switch exactly copies the data frames received or sent by the CPU to a port.

Flow mirror function means that the switch exactly copies the data frames received or by the specified rule of a port to another port. The flow mirror will take effect only the specified rule is permit.

A chassis switch supports at most 4 mirror destination ports, each boardcard allows a source or destination port of a mirror session. At present, each box switch can set many mirror sessions. There is no limitation on mirror source ports, one port or several ports is allowed. When there are more than one source ports, they can be in the same VLAN or in different VLAN. The source port and destination port can be in different VLAN.

Notice: box switch can't use CPU's rx mirror and port's tx mirror at the same time.

8.8.2 Mirror Configuration Task List

1. Specify mirror destination port
2. Specify mirror source port
1. Specify mirror destination port
2. Specify mirror source port (CPU)
3. Specify flow mirror source

1. Specify mirror destination port

Command	Explanation
Global mode	
monitor session <*session*> destination interface <*interface-number*> **no monitor session <*session*> destination interface <*interface-number*>**	Specifies mirror destination port; the no command deletes mirror destination source port.

2. Specify mirror source port（CPU）

Command	Explanation
Global mode	
monitor session <*session*> source {interface <*interface-list*> | cpu [slot <slotnum>]} {rx| tx| both} **no monitor session <session> source**	Specifies mirror source port; the no command deletes mirror source port.

{interface *<interface-list>* / cpu [slot *<slotnum>*]}	

3. Specify flow mirror source

Command	Explanation
Global mode	
monitor session *<session>* source {interface *<interface-list>*} access-group *<num>* {rx|tx|both} **no monitor session *<session>* source {interface *<interface-list>*} access-group *<num>***	Specifies flow mirror source port and apply rule; the no command deletes flow mirror source port.

8.8.3 Mirror Examples

Example:
The requirement of the configurations is shown as below: to monitor at interface 1 the data frames sent out by interface 9 and received from interface 7, sent and received by CPU, and the data frames received by interface 15 and matched by rule 120(The source IP address is 1.2.3.4 and the destination IP address is 5.6.7.8).

Configuration guidelines:
1. Configure interface 1 to be a mirror destination interface.
2. Configure the interface 7 ingress and interface 9 egress to be mirrored source.
3. Configure the CPU as one of the source.
4. Configure access list 120.
5. Configure access 120 to binding interface 15 ingress.

Configuration procedure is as follows:

Switch(config)#monitor session 4 destination interface ethernet 1/0/1
Switch(config)#monitor session 4 source interface ethernet 1/0/7 rx
Switch(config)#monitor session 4 source interface ethernet 1/0/9 tx
Switch(config)#monitor session 4 source cpu
Switch(config)#access-list 120 permit tcp 1.2.3.4 0.0.0.255 5.6.7.8 0.0.0.255
Switch(config)#monitor session 4 source interface ethernet 1/0/15 access-list 120 rx

8.8.4 Device Mirror Troubleshooting

If problems occur on configuring port mirroring, please check the following first for causes:

- Whether the mirror destination port is a member of a TRUNK group or not, if yes, modify the TRUNK group.
- If the throughput of mirror destination port is smaller than the total throughput of

mirror source port(s), the destination port will not be able to duplicate all source port traffic; please decrease the number of source ports, duplicate traffic for one direction only or choose a port with greater throughput as the destination port. Mirror destination port can not be pulled into Isolate vlan, or will affect mirror between VLAN.

When sending unknown unicast, broadcast and multicast data, it is recommended to set only one session for egress mirror.

8.9 RSPAN

8.9.1 Introduction to RSPAN

Port mirroring refers to the duplication of data frames sent/received on a port to another port. The duplicated port is referred to as mirror source port and the duplicating port is referred to as mirror destination port. It is more convenience for network administrator to monitor and manage the network and diagnostic after the mirroring function achieved. But it only used for such instance that the mirror source port and the mirror destination ports are located in the same switch.

RSPAN (remote switched port analyzer) refers to remote port mirroring. It eliminates the limitation that the source port and the destination port must be located on the same switch. This feature makes it possible for the source port and the destination port to be located on different devices in the network, and facilitates the network administrator to manage remote switches. It can't forward traffic flows on remote mirror VLAN.

There are three types of switches with the RSPAN enabled:

1. Source switch: The switch to which the monitored port belongs. The source switch copies the mirrored traffic flows to the Remote VLAN, and then through Layer 2 forwarding, the mirrored flows are sent to an intermediate switch or destination switch.
2. Intermediate switch: Switches between the source switch and destination switch on the network. Intermediate switch forwards mirrored flows to the next intermediate switch or the destination switch. Circumstances can occur where no intermediate switch is present, if a direct connection exists between the source and destination switches.
3. Destination switch: The switch to which the destination port for remote mirroring belongs. It forwards mirrored flows it received from the Remote VLAN to the monitoring device through the destination port.

When configuring the RSPAN mirroring of the source switch, reflector port mode or destination mirror port mode can be selected. The destination switch will redirect all the data frames in the RSPAN VLAN to the RSPAN destination port. For RSPAN mirroring, normal mode and advanced mode can be chosen, normal is introduced by default and fit the normal user. The

advanced mode fit the advanced user.

1. Advanced mode: To redirect data frames in RSPAN VLAN to the RSPAN destination port, the intermediary and destination devices should support the redirection of flow.
2. Normal mode: To configure the RSPAN destination port in the RSPAN VLAN. Thus, datagrams in the RSPAN VLAN will be broadcasted to the destination port. In this mode, the destination port should be in RSPAN VLAN, and the source port should not be configured for broadcasting storm control. TRUNK ports should be configured carefully in order not to forward RSPAN datagrams to external networks. The normal mode has the benefit of easy configuration, and reduced system resources.

To be noticed: Normal mode is introduced by default. When using the normal mode, datagrams with reserved MAC addresses cannot be broadcasted.

For chassis switches, at most 4 mirror destination ports are supported, and source or destination port of one mirror session can be configured on each line card. For box switches, only one mirror session can be configured.The number of the source mirror ports is not limited, and can be one or more. Multiple source ports are not restricted to be in the same VLAN. The destination port and the source ports can be in different VLAN.

For configuration of RSPAN, a dedicated RSPAN VLAN should be configured first for carrying the RSPAN datagrams. The default VLAN, dynamic VLAN, private VLAN, multicast VLAN, and the layer 3 interface enabled VLAN cannot be configured as the RSPAN VLAN. The reflector port must belong to the RSPAN VLAN. The destination port should be connected to the Monitor and the configured as access port or the TRUNK port. The RSPAN reflector port will be working dedicatedly for mirroring, when a port is configured as a reflector port, it will discards all the existing connections to the remote peer, disable configurations related to loopback interfaces, and stop forwarding datagram. Connectivity between the source and destination switch for Remote VLAN, should be made sure by configuration.

To be noticed:

1. Layer 3 interfaces related to RSPAN VLAN should not be configured on the source, intermediate, and the destination switches, or the mirrored datagrams may be discarded.

2. For the source and intermediate switches in the RSPAN connections, the native VLAN of TRUNK port cannot be configured as the RSPAN VLAN, Otherwise the RSPAN tag will be disposed before reaching the destination switches.

3. The source port, in access or trunk mode, should not be added to RSPAN VLAN if advanced RSPAN mode is chosen. When the reflector port is used for a inter-card mirroring of CPU TX data, it must be configured as TRUNK port and allows the RSPAN VLAN data passing, the Native VLAN should not be configured as RSPAN VLAN.

4. When configuring the remote mirroring function, the network bandwidth should be considered in order to carry the network flow and the mirrored flow.

Keywords:

RSPAN: Remote Switched Port Analyzer.

RSPAN VLAN: Dedicated VLAN for RSPAN.

RSPAN Tag: The VLAN tag which is attached to MTP of the RSPAN datagrams.

Reflector Port: The local mirroring port between the RSPAN source and destination ports,

which is not directly connected to the intermediate switches.

8.9.2 RSPAN Configuration Task List

1. Configure RSPAN VLAN
2. Configure mirror source port（cpu）
3. Configure mirror destination port
4. Configure reflector port
5. Configure remote VLAN of mirror group

1. Configure RSPAN VLAN

Command	Explanation
VLAN Configuration Mode	
remote-span **no remote-span**	To configure the specified VLAN as RSPAN VLAN. The no command will remove the configuration of RSPAN VLAN.

2. Configure mirror source port(CPU)

Command	Explanation
Global Mode	
**monitor session *<session>* source {interface *<interface-list>*	cpu [slot *<slotnum>*]} {rx| tx| both}** **no monitor session *<session>* source {interface *<interface-list>*	cpu [slot *<slotnum>*]}**	To configure mirror source port; The no command deletes the mirror source port.

3. Configure mirror destination port

Command	Explanation
Global Mode	
monitor session *<session>* destination interface *<interface-number>* **no monitor session *<session>* destination interface *<interface-number>***	To configure mirror destination interface; The no command deletes the mirror destination port.

4. Configure reflector port

Command	Explanation
Global Mode	
monitor session <session> reflector-port <interface-number> **no monitor session <session> reflector-port**	To configure the interface to reflector port; The no command deletes the reflector port.

5. Configure remote VLAN of mirror group

Command	Explanation
Global Mode	
monitor session *<session>* **remote vlan *<vid>*** **no monitor session <session> remote vlan**	To configure remote VLAN of mirror group, the no command deletes the remote VLAN of mirror group.

8.9.3 Typical Examples of RSPAN

Before RSPAN is invented, network administrators had to connect their PCs directly to the switches, in order to check the statistics of the network.

However, with the help of RSPAN, the network administrators can configure and supervise the switches remotely, which brings more efficiency. The figure below shows a sample application of RSPAN.

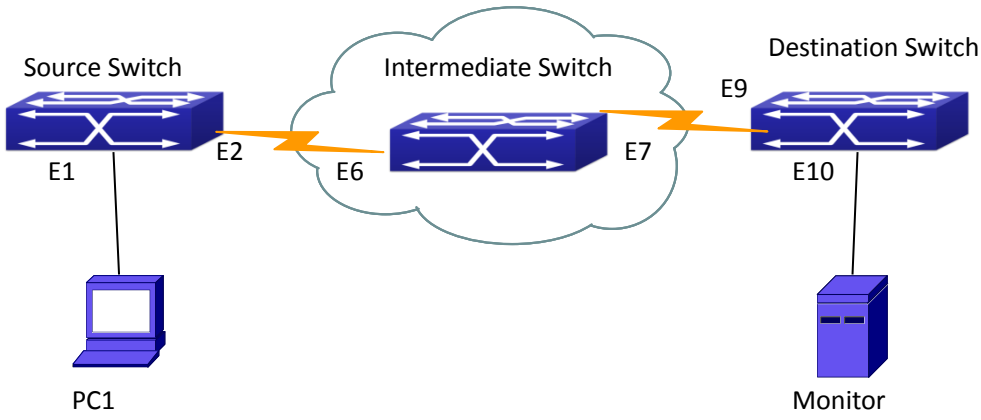


Figure 8-2 RSPAN Application Sample

Two configuration solutions can be chosen for RSPAN: the first is without reflector port, and the other is with reflector port. For the first one, only one fixed port can be connected to the intermediate switch. However, no reflector port has to be configured. This maximizes the usage of witch ports. For the latter one, the port connected to the intermediate switch is not fixed. Datagrams can be broadcasted in the RSPAN VLAN through the loopback, which is much more flexible.

The normal mode configuration is show as below:

Solution 1:

Source switch:

Interface ethernet 1/0/1 is the source port for mirroring.

Interface ethernet 1/0/2 is the destination port which is connected to the intermediate switch.

RSPAN VLAN is 5.

Switch(config)#vlan 5

Switch(Config-Vlan5)#remote-span

Switch(Config-Vlan5)#exit

Switch(config)#interface ethernet 1/0/2

Switch(Config-If-Ethernet1/0/2)#switchport mode trunk
Switch(Config-If-Ethernet1/0/2)#exit
Switch(config)#monitor session 1 source interface ethernet1/0/1 rx
Switch(config)#monitor session 1 destination interface ethernet1/0/2
Switch(config)#monitor session 1 remote vlan 5

Intermediate switch:
Interface ethernet1/0/6 is the source port which is connected to the source switch.
Interface ethernet1/0/7 is the destination port which is connected to the intermediate switch. The native VLAN of this port cannot be configured as RSPAN VLAN, or the mirrored data may not be carried by the destination switch.
RSPAN VLAN is 5.

Switch(config)#vlan 5
Switch(Config-Vlan5)#remote-span
Switch(Config-Vlan5)#exit
Switch(config)#interface ethernet 1/0/6-7
Switch(Config-If-Port-Range)#switchport mode trunk
Switch(Config-If-Port-Range)#exit

Destination switch:
Interface ethernet1/0/9 is the source port, which is connected to the source switch.
Interface ethernet1/0/10 is the destination port which is connected to the monitor. This port is required to be configured as an access port, and belong to the RSPAN VLAN.
RSPAN VLAN is 5.

Switch(config)#vlan 5
Switch(Config-Vlan5)#remote-span
Switch(Config-Vlan5)#exit
Switch(config)#interface ethernet 1/0/9
Switch(Config-If-Ethernet1/0/9)#switchport mode trunk
Switch(Config-If-Ethernet1/0/9)#exit
Switch(config)#interface ethernet 1/0/10
Switch(Config-If-Ethernet1/0/10)#switchport access vlan 5
Switch(Config-If-Ethernet1/0/10)#exit

Solution 2:
Source switch:
Interface ethernet 1/0/1 is the source port.
Interface ethernet 1/0/2 is the TRUNK port, which is connected to the intermediate switch. The native VLAN should not be a RSPAN VLAN.
Interface Ethernet 1/0/3 is a reflector port. The reflector port belongs the RSPAN VLAN, it is access port or TRUNK port of the RSPAN VLAN.
RSPAN VLAN is 5.

```
Switch(config)#vlan 5
Switch(Config-Vlan5)#remote-span
Switch(Config-Vlan5)#exit
Switch(config)#interface ethernet1/0/2
Switch(Config-If-Ethernet1/0/2)#switchport mode trunk
Switch(Config-If-Ethernet1/0/2)#exit
Switch(config)#interface ethernet 1/0/3
Switch(Config-If-Ethernet1/0/3)#switchport mode trunk
Switch(Config-If-Ethernet1/0/3)#exit
Switch(config)#monitor session 1 source interface ethernet1/0/1 rx
Switch(config)#monitor session 1 reflector-port ethernet1/0/3
Switch(config)#monitor session 1 remote vlan 5
```

Intermediate switch:

Interface ethernet1/0/6 is the source port which is connected to the source switch.

Interface ethernet1/0/7 is the destination port which is connected to the destination switch. The native VLAN of the port should not be configured as RSPAN VLAN, or the mirrored data may not be carried by the destination switch.

RSPAN VLAN is 5.

```
Switch(config)#vlan 5
Switch(Config-Vlan5)#remote-span
Switch(Config-Vlan5)#exit
Switch(config)#interface ethernet 1/0/6-7
Switch(Config-If-Port-Range)#switchport mode trunk
Switch(Config-If-Port-Range)#exit
```

Destination switch:

Interface ethernet1/0/9 is the source port which is connected to the source switch.

Interface ethernet1/0/10 is the destination port which is connected to the monitor. This port is required to be configured as an access port, and belong to the RSPAN VLAN.

RSPAN VLAN is 5.

```
Switch(config)#vlan 5
Switch(Config-Vlan5)#remote-span
Switch(Config-Vlan5)#exit
Switch(config)#interface ethernet 1/0/9
Switch(Config-If-Ethernet1/0/9)#switchport mode trunk
Switch(Config-If-Ethernet1/0/9)#exit
Switch(config)#interface ethernet 1/0/10
Switch(Config-If-Ethernet1/0/10)#switchport access vlan 5
Switch(Config-If-Ethernet1/0/10)#exit
```

8.9.4 RSPAN Troubleshooting

Due to the following reasons, RSPAN may not function:

- ☞ Whether the destination mirror port is a member of the Port-channel group. If so, please change the Port-channel group configuration;
- ☞ The throughput the destination port is less than the total throughput of the source mirror ports. If so, the destination cannot catch all the datagrams from every source ports. To solve the problem, please reduce the number of the source ports, or mirror only single direction data flow, or choose some other port with higher capacity as the destination port.
- ☞ Between the source switch and the intermediate switch, whether the native VLAN of the TRUNK ports is configured as RSPAN VLAN. If so, please change the native VLAN for the TRUNK ports.
- ☞ After configured RSPAN, the vlan tag will be added on the packet of the egress mirror. It will cause the abort error frame on the reflection port, so the default MTU value of the switch should be modified.

8.10 ERSPAN

8.10.1 Introduction to ERSPAN

ERSPAN（Encapsulated Remote Switched Port Analyzer） eliminates the limitation that the source port and the destination port must be located on the same switch. This feature makes it possible for the source port and the destination port to be located on different devices in the network, and facilitates the network administrator to manage remote switches. Compared with the traditional RSPAN, ERSPAN configuration is simpler and it makes the monitored traffic to be transmitted in the specified tunnel.

To be noticed

1. Monitor source of ERSPAN monitor only supports port monitor, it does not support CPU monitor and flow monitor presently.
2. For the source and destination switches in the ERSPAN connections, a tunnel must be existed.
3. When configuring the remote mirror function, the network bandwidth should be considered in order to carry the network flow and the mirrored flow.

Keywards:

ERSPAN: Encapsulated Remote Switched Port Analyzer.

8.10.2 ERSPAN Configuration Task List

1. Specify mirror source port

Command	Explanation
Global Mode	
monitor session *<session>* source {interface *<interface-list>*} {rx| tx| both} **no monitor session *<session>* source {interface *<interface-list>*}**	Specify the mirror source port; the no command deletes the mirror source port.

2. Specify mirror destination tunnel

Command	Explanation
Global Mode	
monitor session *<session>* destination tunnel *<tunnel-number>* **no monitor session *<session>* destination tunnel *<tunnel-number>***	Specify the mirror destination tunnel; the no command deletes the mirror destination tunnel.

8.10.3 Typical Examples of ERSPAN

Before ERSPAN is invented, network administrators had to connect their PCs directly to the switches, in order to check the status of the network.

However, with the help of ERSPAN, network administrators can configure and supervise the switches remotely, which brings more efficiency.

In Layer 3 network, Device A connects to the marketing department through Ethernet 1/0/1, and connects to Ethernet1/0/2 of Device B through Ethernet 1/0/2; Device C connects to the server through Ethernet 1/0/2, and connects to Ethernet 1/0/1 of Device B through Ethernet 1/0/1. Server is able to monitor the bidirectional traffic of the marketing department across a GRE tunnel by configuring remote port mirroring. The figure below shows a sample application of ERSPAN.

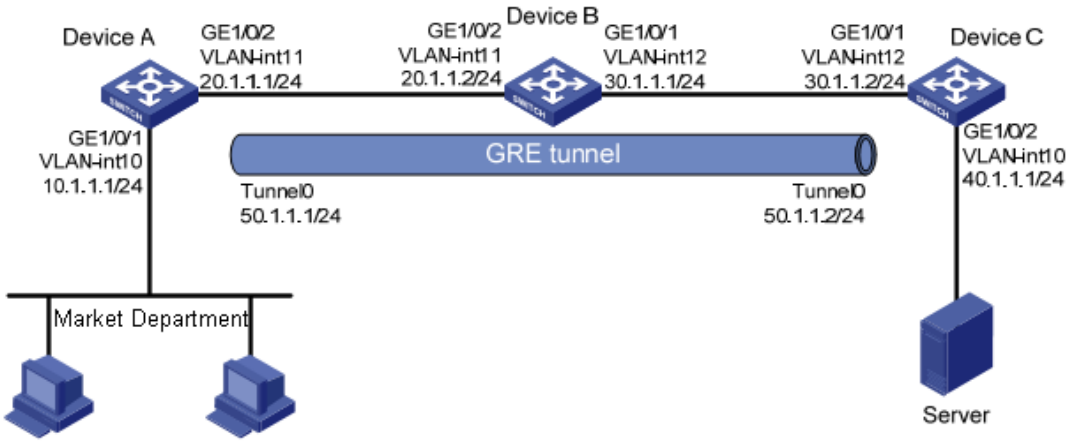


Figure 8-3 ERSPAN application diagram

Before configuring layer-3 remote port mirroring, make sure that you have created a GRE tunnel that connects the source and destination device, and ensure the normal transmitting for GRE tunnel.

The configuration of layer-3 remote port mirror needs to be processed on the source and destination devices respectively. Both the source and destination ports are configured on the source and destination devices, the difference in the following:

1) On Device A, configure the port which you want to monitor as the source port and configure the tunnel interface as the destination port.

2) On Device C, configure the physical port corresponding to the tunnel interface as the source port and configure the port that connects the data monitor device as the destination port.

(1) Configure IP addresses

Configure IP address and subnet mask for the interfaces, configuration procedures is omitted.

(2) Configure Device A (the source device)

Create interface Tunnel1, and configure an IP address and mask for it.

SwitchA(config)#interface tunnel 1

SwitchA(config-if-tunnel1)# tunnel mode gre ip

SwitchA (config-if-tunnel1)#ip address 50.1.1.1 255.255.255.0

Configure Tunnel1 to operate in GRE tunnel mode, and configure source and destination IP addresses for it.

SwitchA(config-if-tunnel1)# tunnel source 10.1.1.1

SwitchA(config-if-tunnel1)# tunnel destination 40.1.1.1

SwitchA(config-if-tunnel1)# exit

Configure OSPF protocol.

SwitchA (config)#router ospf

SwitchA (config-router)#network 0.0.0.0/0 area 0

SwitchA (config-router)#exit

Configure Ethernet 1/0/1 as a source port and Tunnel1 as the destination port of local mirroring group 1.

SwitchA(config)#monitor session 4 destination tunnel 1

SwitchA(config)#monitor session 4 source interface ethernet 1/0/1 both

(3) Configure Device B (the intermediate device)

Configure OSPF protocol.

SwitchB (config)#router ospf

SwitchB (config-router)#network 0.0.0.0/0 area 0

SwitchB (config-router)#exit

(4) Configure Device C (the destination device)

Create interface Tunne1, and configure an IP address and mask for it.

SwitchC(config)#interface tunnel 1

SwitchC (config-if-tunnel1)# tunnel mode gre ip

SwitchC (config-if-tunnel1)#ip address 50.1.1.2 255.255.255.0

Configure Tunnel1 to operate in GRE tunnel mode, and configure source and destination IP addresses for it.

SwitchC (config-if-tunnel1)# tunnel source 40.1.1.1

SwitchC (config-if-tunnel1)# tunnel destination 10.1.1.1
SwitchC (config-if-tunnel1)# exit
Configure OSPF protocol.
SwitchC (config)#router ospf
SwitchC (config-router)#network 0.0.0.0/0 area 0
SwitchC (config-router)#exit
Configure Ethernet 1/0/1 as a source port and Ethernet 1/0/2 as the destination port of local mirroring group 1.
SwitchC (config)#monitor session 1 destination interface ethernet 1/0/2
SwitchC (config)#monitor session 1 source interface ethernet 1/0/1 rx

8.10.4 ERSPAN Troubleshooting

If problems occur when configuring ERSPAN, please check whether the problem is caused by the following reasons:

- Make sure GRE tunnel configuration to ensure the normal transmission for the traffic.
- If the throughput of mirror destination port is smaller than the total throughput of mirror source port(s), the destination port will not be able to duplicate the traffic of all source port; please decrease the number of source ports, duplicate traffic for one direction only or choose a port with greater throughput as the destination port.

8.11 sFlow

8.11.1 Introduction to sFlow

The sFlow (RFC 3176) is a protocol based on standard network export and used on monitoring the network traffic information developed by the InMon Company. The monitored switch or router sends date to the client analyzer through its main operations such as sampling and statistic, then the analyzer will analyze according to the user requirements so to monitor the network.

A sFlow monitor system includes: sFlow proxy, central data collector and sFlow analyzer. The sFlow proxy collects data from the switch using sampling technology. The sFlow collector is for formatting the sample data statistic which is to be forwarded to the sFlow analyzer which will analyze the sample data and perform corresponding measure according to the result. Our switch here acts as the proxy and central data collector in the sFlow system.

We have achieved data sampling and statistic targeting physical port.

Our data sample includes the IPv4 and IPv6 packets. Extensions of other types are not supported so far. As for non IPv4 and IPv6 packet, the unify HEADER mode will be adopted

following the requirements in RFC3176, copying the head information of the packet based on analyzing the type of its protocol.

The latest sFlow protocol presented by InMon Company is the version 5. Most of the sFlow analysis software has support for version 5.The message format defined in version 5 differs slightly from that in version 4.Our implementation is already compatible with that part of the content.

8.11.2 sFlow Configuration Task List

1. Configure sFlow Collector address

Command	Explanation
Global mode and Port Mode	
sflow destination *<collector-address>* [<*collector-port*>] **no sflow destination**	Configure the IP address and port number of the host in which the sFlow analysis software is installed. As for the ports, if IP address is configured on the port, the port configuration will be applied, or else will be applied the global configuration. The "**no sflow destination**" command restores to the default port value and deletes the IP address.

2. Configure the sFlow proxy address

Command	Explanation
Global Mode	
sflow agent-address *<collector-address>* **no sflow agent-address**	Configure the source IP address applied by the sFlow proxy; the "no" form of the command deletes this address.

3. Configure the sFlow proxy priority

Command	Explanation
Global Mode	
sflow priority *<priority-vlaue>* **no sflow priority**	Configure the priority when sFlow receives packet from the hardware; the "**no sflow priority**" command restores to the default

4. Configure the packet head length copied by sFlow

Command	Explanation
Port Mode	
sflow header-len *<length-vlaue>* **no sflow header-len**	Configure the length of the packet data head copied in the sFlow data sampling; the "no" form of this command restores to the default value.

5. Configure the max data head length of the sFlow packet

Command	Explanation
Port Mode	

sflow data-len *<length-vlaue>* **no sflow data-len**	Configure the max length of the data packet in sFlow; the “no” form of this command restores to the default.

6. Configure the sampling rate value

Command	Explanation
Port Mode	
sflow rate {input *<input-rate>* | output *<output-rate >*} **no sflow rate [input | output]**	Configure the sampling rate when sFlow performing hardware sampling. The “no” command deletes the rate value.

7. Configure the sFlow statistic sampling interval

Command	Explanation
Port Mode	
sflow counter-interval *<interval-vlaue>* **no sflow counter-interval**	Configure the max interval when sFlow performing statistic sampling. The “no” form of this command deletes

8. Configure the analyzer used by sFlow

Command	Explanation
Global Mode	
sflow analyzer sflowtrend **no sflow analyzer sflowtrend**	Configure the analyzer used by sFlow, the no command deletes the analyzer.

9. Configure the global sflow version

Command	Explanation
Global Mode	
sflow version<version> **no sflow version**	Configure the global sflow version. The no command recovers to be the default value. Default version is 4.

8.11.3 sFlow Examples



Figure 8-4 sFlow configuration topology

As shown in the figure, sFlow sampling is enabled on the port 1/0/1 and 1/0/2 of the switch. Assume the sFlow analysis software is installed on the PC with the address of 192.168.1.200. The address of the layer 3 interface on the SwitchA connected with PC is 192.168.1.100. A loopback interface with the address of 10.1.144.2 is configured on the SwitchA. sFlow configuration is as follows:

Configuration procedure is as follows:

```
Switch#config
Switch (config)#sflow ageng-address 10.1.144.2
Switch (config)#sflow destination 192.168.1.200
Switch (config)#sflow priority 1
Switch (config)# interface ethernet1/0/1
Switch (Config-If-Ethernet1/0/1)#sflow rate input 10000
Switch (Config-If-Ethernet1/0/1)#sflow rate output 10000
Switch (Config-If-Ethernet1/0/1)#sflow counter-interval 20
Switch (Config-If-Ethernet1/0/1)#exit
Switch (config)# interface ethernet1/0/2
Switch (Config-If-Ethernet1/0/2)#sflow rate input 20000
Switch (Config-If-Ethernet1/0/2)#sflow rate output 20000
Switch (Config-If-Ethernet1/0/2)#sflow counter-interval 40
```

8.11.4 sFlow Troubleshooting

In configuring and using sFlow, the sFlow server may fail to run properly due to physical connection failure, wrong configuration, etc. The user should ensure the following:

- ☞ Ensure the physical connection is correct
- ☞ Guarantee the address of the sFlow analyzer configured under global or port mode is accessible.
- ☞ If traffic sampling is required, the sampling rate of the interface must be configured
- ☞ If statistic sampling is required, the statistic sampling interval of the interface must be configured
- ☞ If the sFlow version is the same as the version of sFlow analysis software used
- ☞ If the examination remains unsolved, please contact with the technical service center of our company.

Chapter 9 Network Time Management Configuration

9.1 NTP

9.1.1 Introduction to NTP Function

The NTP (Network Time Protocol) synchronizes timekeeping spans WAN and LAN among distributed time servers and clients, it can get millisecond precision. The introduction of event, state, transmit function and action are defined in RFC-1305.

The purpose of using NTP is to keep consistent timekeeping among all clock-dependent devices within the network so that the devices can provide diverse applications based on the consistent time.

For a local system running NTP, its time can be synchronized by other reference sources and can be used as a reference source to synchronize other clocks, also can synchronize each other by transmit NTP packets.

9.1.2 NTP Function Configuration Task List

1. To enable NTP function
2. To configure NTP server function
3. To configure the max number of broadcast or multicast servers supported by the NTP client
4. To configure time zone
5. To configure NTP access control list
6. To configure NTP authentication
7. To specified some interface as NTP broadcast/multicast client interface
8. To configure some interface can’t receive NTP packets
9. To configure the request packet sending interval of ntp client
10. Display information
11. Debug

1. To enable NTP function

Command	Explication
Global Mode	
ntp enable **ntp disable**	To enable or disable NTP function.

2. To configure NTP server function

Command	Explication

Global Mode	
ntp server {<*ip-address*> / <*ipv6-address*>} [version <*version_no*>] [key <*key-id*>] **no ntp server {<*ip-address*> / <*ipv6-address*>}**	To enable the specified time server of time source.

3. To configure the max number of broadcast or multicast servers supported by the NTP client

Command	Explication
Global Mode	
ntp broadcast server count <number> **no ntp broadcast server count**	Set the max number of broadcast or multicast servers supported by the NTP client. The no operation will cancel the configuration and restore the default value.

4. To configure time zone

Command	Explication
Global Mode	
clock timezone WORD {add | subtract} <0-23> [<0-59>] **no clock timezone WORD**	This command configures timezone in global mode, the no command deletes the configured timezone.

5. To configure NTP access control list

Command	Explication
Global Mode	
ntp access-group server <*acl*> **no ntp access-group server < *acl*>**	To configure NTP server access control list.

6. To configure NTP authentication

Command	Explication
Global Mode	
ntp authenticate **no ntp authenticate**	To enable NTP authentication function.
ntp authentication-key <*key-id*> md5 <*value*> **no ntp authentication-key <*key-id*>**	To configure authentication key for NTP authentication.
ntp trusted-key <*key-id*> **no ntp trusted-key <*key-id*>**	To configure trusted key.

7. To specified some interface as NTP broadcast/multicast client interface

Command	Explication
vlan Configuration Mode	

ntp broadcast client **no ntp broadcast client**	To configure specified interface to receive NTP broadcast packets.
ntp multicast client **no ntp multicast client**	To configure specified interface to receive NTP multicast packets.
ntp ipv6 multicast client **no ntp ipv6 multicast client**	To configure specified interface to receive IPv6 NTP multicast packets.

8. To configure some interface can't receive NTP packets

Command	Explication
vlan Configuration Mode	
ntp disable **no ntp disable**	To disable the NTP function.

9. To configure the request packet sending interval of ntp client

Command	Explication
Global Mode	
ntp syn-interval <1-3600> **no ntp syn-interval**	Configure the request packet sending interval of ntp client as 1s-3600s. The no command recovers to be the default value of 64s.

10. Display information

Command	Explication
Admin Mode	
show ntp status	To display the state of time synchronize.
show ntp session [<*ip-address*> | <*ipv6-address*>]	To display the information of NTP session.

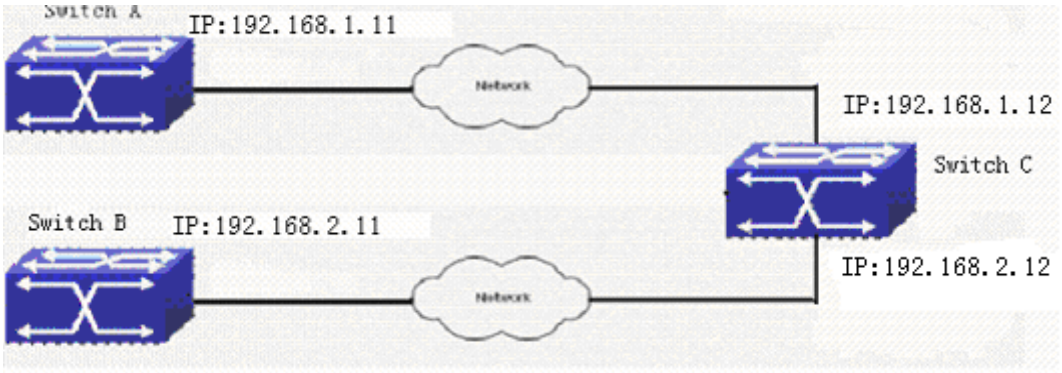
11. Debug

Command	Explication
Admin Mode	
debug ntp authentication **no debug ntp authentication**	To enable debug switch of NTP authentication.
debug ntp packet [send | receive] **no debug ntp packet [send | receive]**	To enable debug switch of NTP packet information.

debug ntp adjust **no debug ntp adjust**	To enable debug switch of time update information.
debug ntp sync **no debug ntp sync**	To enable debug switch of time synchronize information.
debug ntp events **no debug ntp events**	To enable debug switch of NTP event information.

9.1.3 Typical Examples of NTP Function

A client switch wanted to synchronize time with time server in network, there is two time server in network, the one is used as host, the other is used as standby, the connection and configuration as follows (Switch A and Switch B are the switch or route which support NTP server):



The configuration of Switch C is as follows: (Switch A and Switch B may have the different command because of different companies, we not explain there, our switches are not support NTP server at present)

Switch C:

Switch(config)#ntp enable

Switch(config)#interface vlan 1

Switch(Config-if-Vlan1)#ip address 192.168.1.12 255.255.255.0

Switch(config)#interface vlan 2

Switch(Config-if-Vlan1)#ip address 192.168.2.12 255.255.255.0

Switch(config)#ntp server 192.168.1.11

Switch(config)#ntp server 192.168.2.11

9.1.4 NTP Function Troubleshooting

In configuration procedures, if there is error occurred, the system can give out the debug

information.

The NTP function disables by default, the show command can be used to display current configuration. If the configuration is right please use debug every relative debugging command and display specific information in procedure, and the function is configured right or not, you can also use show command to display the NTP running information, any questions please send the recorded message to the technical service center.

9.2 SNTP

9.2.1 Introduction to SNTP

The Network Time Protocol (NTP) is widely used for clock synchronization for global computers connected to the Internet. NTP can assess packet sending/receiving delay in the network, and estimate the computer's clock deviation independently, so as to achieve high accuracy in network computer clocking. In most positions, NTP can provide accuracy from 1 to 50ms according to the characteristics of the synchronization source and network route.

Simple Network Time Protocol (SNTP) is the simplified version of NTP, removing the complex algorithm of NTP. SNTP is used for hosts who do not require full NTP functions; it is a subset of NTP. It is common practice to synchronize the clocks of several hosts in local area network with other NTP hosts through the Internet, and use those hosts to provide time synchronization service for other clients in LAN. The figure below depicts a NTP/SNTP application network topology, where SNTP mainly works between second level servers and various terminals since such scenarios do not require very high time accuracy, and the accuracy of SNTP (1 to 50 ms) is usually sufficient for those services.

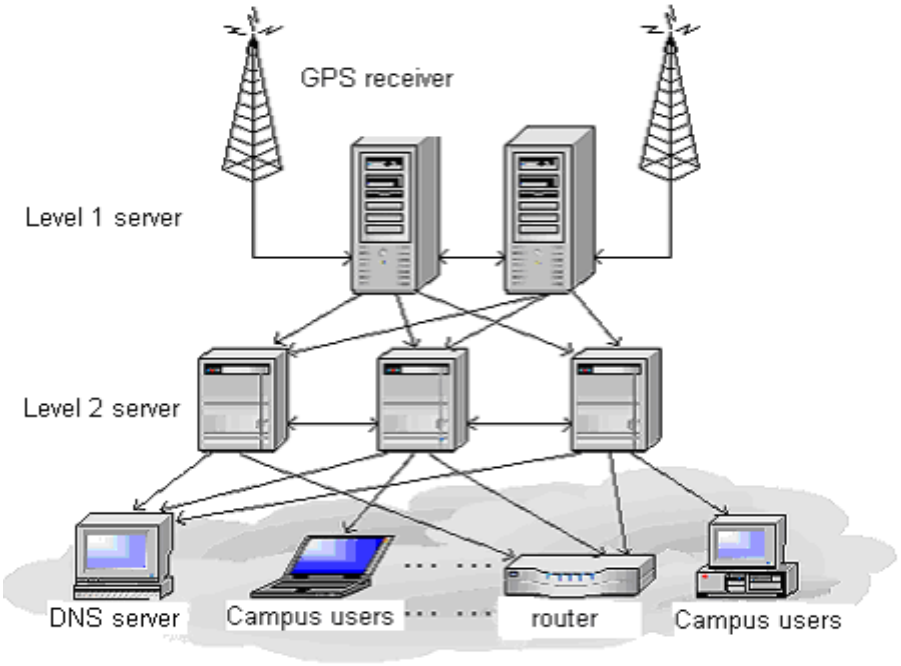


Figure 9-1 Working Scenario

Switch implements SNTPv4 and supports SNTP client unicast as described in RFC2030; SNTP client multicast and unicast are not supported, nor is the SNTP server function.

9.2.2 Typical Examples of SNTP Configuration

/

Figure 9-2 Typical SNTP Configu

All switches in the autonomous zone are required to perfo e synchronization, which done through two redundant SNTP/NTP servers. For time to be synchronized, the be properly configured. There should be reachable route between any switc SNTP/NTP servers.

SNTP/NTP SERVER

SWITCH SWITCH

Example: Assume the IP addresses of the SNTP/NTP servers are 10.1. respectively, and SNTP/NTP server function (such as NTP master) is enabled, the for any switch should like the following:

Switch#config

Switch(config)#sntp server 10.1.1.1

9.3 DNSv4/v6

9.3.1 Introduction to DNS

DNS (Domain Name System) is a distributed database used by TCP/IP applications to

translate domain names into corresponding IPv4/IPv6 addresses. With DNS, you can use easy-to-remember and signification domain names in some applications and let the DNS server translate them into correct IPv4/IPv6 addresses.

There are two types of DNS services, static and dynamic, which supplement each other in application. Each time the DNS server receives a name query it checks its static DNS database first before looking up the dynamic DNS database. Some frequently used addresses can be put in the static DNS database, the reduction the searching time in the dynamic DNS database would increase efficiency. The static domain name resolution means setting up mappings between domain names and IPv4/IPv6 addresses. IPv4/IPv6 addresses of the corresponding domain names can be found in the static DNS database when you use some applications. Dynamic domain name resolution is implemented by querying the DNS server. A user program sends a name query to the resolver in the DNS client when users want to use some applications with domain name, the DNS resolver looks up the local domain name cache for a match. If a match is found, it sends the corresponding IPv4/IPv6 address back to the switch. If no match is found, it sends a query to a higher DNS server. This process continues until a result, whether success or failure, is returned.

The Domain Name System (DNS) is a hierarchical naming system for computers, services, or any resource participating in the Internet. It associates various information with domain names assigned to such participants. Most importantly, it translates humanly meaningful domain names to the numerical (binary) identifiers associated with networking equipment for the purpose of locating and addressing these devices world-wide. An often used analogy to explain the Domain Name System is that it serves as the "phone book" for the Internet by translating human-friendly computer hostnames into IP addresses.

The Domain Name System makes it possible to assign domain names to groups of Internet users in a meaningful way, independent of each user's physical location. Because of this, World-Wide Web (WWW) hyperlinks and Internet contact information can remain consistent and constant even if the current Internet routing arrangements change or the participant uses a mobile device. Internet domain names are easier to remember than IP addresses such as 208.77.188.166(IPv4) or 2001:db8:1f70::999:de8:7648:6e8 (IPv6). People take advantage of this when they recite meaningful URLs and e-mail addresses without having to know how the machine will actually locate them.

The Domain Name System distributes the responsibility for assigning domain names and mapping them to Internet Protocol (IP) networks by designating authoritative name servers for each domain to keep track of their own changes, avoiding the need for a central register to be continually consulted and updated.

In general, the Domain Name System also stores other types of information, such as the list of mail servers that accept email for a given Internet domain. By providing a world-wide, distributed keyword-based redirection service, the Domain Name System is an essential component of the functionality of the Internet.

9.3.2 DNSv4/v6 Configuration Task List

1. To enable/disable DNS function
2. To configure/delete DNS server
3. To configure/delete domain name suffix

4. To delete the domain entry of specified address in dynamic cache
5. To enable DNS dynamic domain name resolution
6. Enable/disable DNS SERVER function
7. Configure the max number of client information in the switch queue
8. Configure the timeout value of caching the client information on the switch
9. Monitor and diagnosis of DNS function

1. To enable/disable DNS function

Command	Explanation
Global Mode	
ip domain-lookup **no ip domain-lookup**	To enable/disable DNS dynamic lookup function.

2. To configure/delete DNS server

Command	Explanation
Global Mode	
**dns-server {*<ip-address>*	*<ipv6-address>*}** **[priority *<value>*]** **no dns-server {*<ip-address>*	*<ipv6-address>*}**	To configure DNS server, the no form of this command deletes DNS server.

3. To configure/delete domain name suffix

Command	Explanation
Global Mode	
ip domain-list *<WORD>* **no ip domain-list *<WORD>***	To configure/delete domain name suffix.

4. To delete the domain entry of specified address in dynamic cache

Command	Explanation
Admin Mode	
**clear dynamic-host {*<ip-address>*	*<ipv6-address>*	all}**	To delete the domain entry of specified address in dynamic cache.

5. To enable DNS dynamic domain name resolution

Command	Explanation
Global Mode	
**dns lookup {ipv4	ipv6} *<hostname>***	To enable DNS dynamic domain name resolution.

6. Enable/disable DNS SERVER function

Command	Explanation
Global Mode	

ip dns server **no ip dns server**	Enable/disable DNS SERVER function.

7. Configure the max number of client information in the switch queue

Command	Explanation
Global Mode	
ip dns server queue maximum <1-5000> **no ip dns server queue maximum**	Configure the max number of client information in the switch queue.

8. Configure the timeout value of caching the client information on the switch

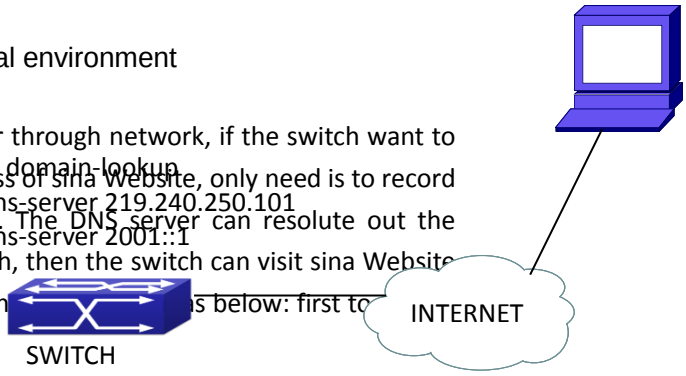
Command	Explanation
Global Mode	
ip dns server queue timeout <1-100> **no ip dns server queue timeout**	Configure the timeout value of caching the client information on the switch.

9. Monitor and diagnosis of DNS function

Command	Explanation
Admin Mode and Configuration Mode	
show dns name-server	To show the configured DNS server information.
show dns domain-list	To show the configured DNS domain name suffix information.
show dns hosts	To show the dynamic domain name information of resolved by switch.
show dns config	Display the configured global DNS information on the switch.
show dns client	Display the DNS Client information maintained by the switch.
debug dns {all | packet [send | recv] | events | relay} **no debug dns {all | packet [send | recv] | events | relay}**	To enable/disable DEBUG of DNS function.

9.3.3 Typical Examples of DNS

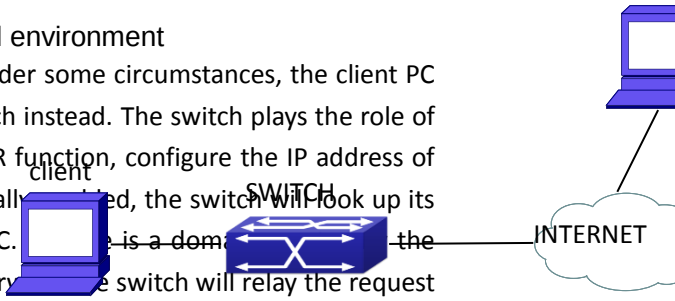
Figure 9-3 DNS CLIENT typical environment



As shown in fig, the switch connected to DNS server through network, if the switch want to visit sina Website, it needn't to know the IPv4/IPv6 address of sina Website, only need is to record the domain name of sina Website is www.sina.com.cn. The DNS server can resolute out the IPv4/IPv6 address of this domain name and send to switch, then the switch can visit sina Website correctly. The switch is configured as DNS client, basic configurations as below: first to

DNS dynamic domain name resolution function on switch, and configure DNS server address, then with some kinds of tools such as PING, the switch can get corresponding IPv4/IPv6 address with dynamic domain name resolution function.

Figure 9-4 DNS SERVER typical environment



The figure above is an application of DNS SERVER. Under some circumstances, the client PC doesn't know the real DNS SERVER, and points to the switch instead. The switch plays the role of a DNS SERVER in two steps: Enable the global DNS SERVER function, configure the IP address of the real DNS server. After the DNS SERVER function is globally enabled, the switch will look up its local cache when receiving a DNS request from a client PC. If the request is a domain name of the local client, it will directly answer the client's request; otherwise the switch will relay the request to the real DNS server, pass the reply from the DNS Server to the client and record the domain and its IP address for a faster lookup in the future.

Switch configuration for DNS CLIENT:
Switch(config)# ip domain-lookup
Switch(config)# dns-server 219.240.250.101
Switch(config)# dns-server 2001::1
Switch#ping host www.sina.com.cn
Switch#traceroute host www.sina.com.cn
Switch#telnet host www.sina.com.cn

Switch configuration for DNS SERVER:
Switch(config)# ip domain-lookup
Switch(config)# dns-server 219.240.250.101
Switch(config)# dns-server 2001::1
Switch(config)# ip dns server

9.3.4 DNS Troubleshooting

In configuring and using DNS, the DNS may fail due to reasons such as physical connection failure or wrong configurations. The user should ensure the following:

- First make sure good condition of the TACACS+ server physical connection;
- Second all interface and link protocols are in the UP state (use "**show interface**" command);
- Then please make sure that the DNS dynamic lookup function is enabled (use the "ip domain-lookup" command) before enabling the DNS CLIENT function. To use DNS SERVER function, please enable it (use the "ip dns server" command);
- Finally ensure configured DNS server address (use "**dns-server**" command), and the switch can ping DNS server;
- If the DNS problems remain unsolved, please use debug DNS all and other debugging command and copy the DEBUG message within 3 minutes, send the recorded message to the technical service center of our company.

9.4 Summer Time

9.4.1 Introduction to Summer Time

Summer time is also called daylight saving time, it is a time system for saving energy sources. In summer the time is advanced 1 hour to keep early hours, reduce the lighting, so as to save electrolighting. The rule that adopt summer time is different in each country. At present, almost 110 countries implement summer time.

Compare with the standard time, usually set summer time 1 hour late, for example, when summer time is implementing, 10:00 am of the standard time is considered 11:00 am of summer time.

9.4.2 Summer Time Configuration Task Sequence

1. Configure absolute or recurrent time range of summer time

Command	Explanation
Global Mode	
clock summer-time <*word*> absolute <*HH:MM*> <*YYYY.MM.DD*> <*HH:MM*> <*YYYY.MM.DD*> [<*offset*>] **no clock summer-time**	Set absolute time range of summer time, start and end summer time is configured with specified year.
clock summer-time <*word*> recurring <*HH:MM*> <*MM.DD*> <*HH:MM*> <*MM.DD*> [<*offset*>] **no clock summer-time**	Set recurrent time range of summer time, every year the summer time begins from the start time and end at the end time.
clock summer-time <*word*> recurring <*HH:MM*> <*week*> <*day*> <*month*> <*HH:MM*> <*week*> <*day*> <*month*> [<*offset*>] **no clock summer-time**	Set recurrent time range of summer time, every year the summer time begins from the start time and end at the end time.

9.4.3 Examples of Summer Time

Example1:

The configuration requirement in the following: The summer time from 23:00 on April 1th, 2012 to 00:00 on October 1th, 2012, clock offset as 1 hour, and summer time is named as 2012.

Configuration procedure is as follows:

Switch(config)# clock summer-time 2012 absolute 23:00 2012.4.1 00:00 2012.10.1

Example2:

The configuration requirement in the following: The summer time from 23:00 on the first Saturday of April to 00:00 on the last Sunday of October year after year, clock offset as 2 hours, and summer time is named as time_travel.

Configuration procedure is as follows:

Switch(config)#clock summer-time time_travel recurring 23:00 first sat apr 00:00 last sun oct 120

9.4.4 Summer Time Troubleshooting

If there is any problem happens when using summer time, please check whether the problem is caused by the following reasons:

- ☞ Check whether command mode in global mode
- ☞ Check whether system clock is correct

Chapter 10 vpn Configuration

10.1 MPLS Overview

10.1.1 MPLS Overview

MPLS (Multiprotocol Label Switching), originating from IPv4, was first designed for improving the forwarding speed. Its core technology can be extended into multiple network protocols, including IPv6 (Internet Protocol version 6), IPX (Internet Packet Exchange), Appletalk, DECnet, CLNP (Connectionless Network Protocol) and etc, since the “Multiprotocol” in MPLS means supporting multiple protocols. MPLS technology is a combination of fast switch and L3 route forwarding hence can satisfy the network requirement of various new applications.

10.1.1.1 MPLS Introduction

Forwarding Equivalence Class

MPLS, as a class-based forwarding technology, will put packets with the same forwarding mode into a class named as FEC (Forwarding Equivalence Class). The same FEC group will be treated with the same way in MPLS networks. FEC is a group of L3 messages, which will be forwarded along the same path, at the same priority level, and in the same mode. There are two steps to finish the forwarding process:

- Analyze the packet header and divide packets into FEC
- Map the FEC to the next-hop

In traditional IP forwarding networks, each router will process the same packet with the above two steps. FEC can include one or more FEC units. All of them are L3 message packets that can be mapped to the same LSP.

At present, there are two types of FEC:

- Address Prefix: Use the Address Prefix to identify a FEC unit, whose length ranges from 0 to the full address length. Each Address Prefix FEC unit corresponds with a destination subnet.
- Host Address: Use the Host Address to identify a FEC unit, as each unit corresponds with a host address.

The division rules of FEC is very flexible, which can be any combination of source address, destination address, source port, destination port, protocol type, VPN and etc. For instance, in the traditional IP forwarding using the Longest Prefix Match Algorism, all packets targeted at the same destination address belong to one FEC.

Label

In MPLS networks,,each specific FEC will be encoded at the edge LSR into a label - a short,

fixed-length value, which will be added to the head of packets and turn them into label packets, before they are forwarded. Besides a segment identifying FEC, labels also include a COS segment, and thus representing FEC, precedence, and service class as a whole. LSR will divide packets reaching different ports into different FEC to establish the foundation of VPN. When a LSR creates a new FEC, it will also create a corresponding label, and advertise it to all peers. LSR maintain both incoming and outgoing labels. To implement load sharing, one FEC may correspond with multiple labels, but one label can only represent one FEC.

Labels, being carried in packet header, don't include topology information, and is only locally meaningful. The label length is 4 bytes. The figure demonstrates its encapsulation structure:



Figure 10-1 The Encapsulation Structure of a Label

There are 4 fields in a label:

- Label：The label value, whose length is 20 bits, a pointer for forwarding.
- Exp：3bits, used by QoS.
- S：1bit, the label's layered structure supported by MPLS, that is, there are multiple label layers. The value 1 represents the bottom-most layer of label.
- TTL：8bits, serves the same purpose as the TTL (Time To Live) in IP packets.

The label, like VPI/VCI of ATM and DLCI of Frame Relay, is identification for connections. If there is a label field in the link-layer protocol, such as VPI/VCI of ATM and DLCI of Frame Relay, the label will be encapsulated in these fields, otherwise, in a transitional layer between the link layer and the IP layer. Thus, labels can be supported by any link layer protocol.

Label Space

LSR can distribute a different label for a FEC according to its ingress port. As a result, packets from different ports can be forwarded independently, which is the basic foundation of VPN. To enhance the utilization efficiency of labels, MPLS provides the concept of label space, which is a label prefix. By allocating FECs belonging to different label spaces with the same label, the boundary of label is actually expanded. The label space is only meaningful when allocating labels, but not when forwarding them.

Label Switching

There is no need to analyze packet header in non-edge LSRs, instead, the label will be used as a pointer to the next-hop egress port and a new label. The label packet will replace the old label with the new one and then be forwarded through the specified egress port. Label switching will simplify and accelerate the forwarding process, and realize applications like VPN, QoS, traffic engineering and etc.

Label Switching Router

LSR (Label Switching Router) is the basic element of a MPLS network, with all LSRs supporting MPLS technology.

LSR is a device able to forwarding packets according to their label value. A LSR connecting an IP route network and a MPLS switching network is called an Edge LSR. Such a LSR is able to adding labels to IP messages and forwarding data according to LSP, or deleting MPLS packet labels and forwarding data according to the IP routes. Each LSR must be distributed a global-alone LSR ID,

usually get an interface IP address of LSR. Assume that, LSR Ru and Rd agree on the map between the label L and the FEC F. Packets can be forwarded from Ru to Rd based on the label L, in which case, Ru is the upstream LSR, and Rd the downstream LSR, that is to say, the forwarding of packets id always from the upstream LSR to the downstream one.

Label Switched Path

The path a FEC follows in the MPLS network is called a LSP (Label Switched Path). Two adjacent LSRs in a LSP are separately called the upstream and downstream LSR, along the direction of data transmission. In the next figure, R2 is the downstream LSR of R1,while R1 is the upstream LSR of R2.

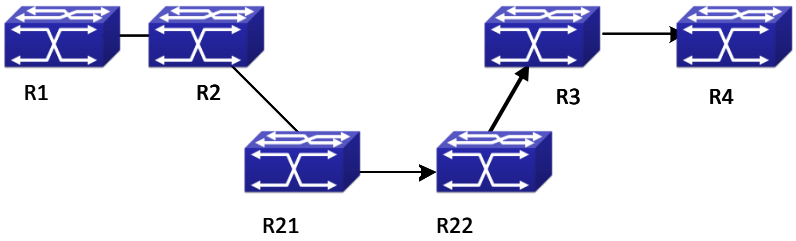


Figure 10-2 Label Switched Path LSP

The function of LSP, the same as the virtual circuit of ATM and Frame Relay, is a unidirectional path form the ingress of a MPLS network to its egress. Each router along the LSP is a LSR.

When downstream LSRs advertise labels to their upstream LSRs, all labels as a series and the LSR sequence compose a LSP. LSP will map the IP layer route information to a link layer switched path. LSP is a unidirectional packet forwarding path, along which, packets are always forwarded form an upstream LSR to a downstream one. To forward packets in the opposite direction, creating an entirely new and independent LSP is necessary. LSP always relates FEC with LSP. This relationship between FEC and LSP is called mapping packets to LSP.

The rules of mapping packets to LSP:

If there is only one LSP, which includes a host-address FEC unit with the same destination address as the packet, map the packet to it;

If there is more than one LSP satisfying condition 1, map the packet to any one of them.

If there is only one LSP, whose address-prefix FEC unit can match the packet, map the packet to it.

If there is more than one LSP satisfying condition 3, choose a LSP based on the Longest Prefix Match principle;

If a packet will definitely pass through a specific egress LSR, and there is a LSP, the prefix FEC unit bounded to which is the address of that egress LSR, map the packet to this LSP.

Additional Rules:

If the destination address of the packet matches no LSP, the packet will be sent along the LSP with the same address as its Egress Router, as long as the LSP has an Address-prefix FEC unit.

If a packet matches two LSPs, one of which includes a host-address FEC unit, and the other an address-prefix FEC unit, always map the packet to the first one.

If the packet matches no LSP with a host-address FEC unit, it should not be sent along a LSP even if whose host-address FEC unit is the same as the packet's egress router address.

The creation of LSP is based on connections, which are the result of topology information rather than the demand of data flow. That is to say, no matter data forwarded by this router exist or not, the LSP will always be created.

Label Merging

With the LSR mapping multiple incoming labels to the same FEC, all these incoming labels will correspond with the same outgoing label and egress port. As a result, when packets with different labels reach the LSR, all outgoing packets will carry the same label. This process is called Label Merging. Label Merging can decrease the label number in the MPSL domain, but maybe at the cost of losing ingress port information of the packets.

If the LSR doesn't support label merging, when there are multiple label requests, it will initiate a new label request to the downstream LSR once for each of them, no matter they have the same FEC or not. Otherwise, only one label request will be implemented.

Label Distribution Protocol

LDP (Label Distribution Protocol) is the MPLS control protocol, like singling protocols in traditional networks, whose function includes classifying FEC, distributing labels, creating and maintaining LSP and etc.

MPSL supports multiple label distribution protocols, including protocols specially designed for distributing labels, like LDP, CR-LDP (Constraint-Based Routing using LDP), and existing ones capable of it after extension, like BGP (Border Gateway Protocol), RSVP (Resource Reservation Protocol). Besides, manually configured static LSP is allowed.

LSP Tunnel Technology

MPLS supports LSP tunnel technology. Even if the path between an upstream LSR and a downstream LSR in a LSP is not provided by the routing protocol, MPLS allows creating a new LSP connecting the two, making them the start and end of it separately. This new LSP is a LSP tunnel, which avoids encapsulating the tunnel via traditional network layer.

If the routes passed by a tunnel are the same as those from the routing protocol, this tunnel is Hop-by-Hop Routed Tunnel; or, it is an Explicitly Routed Tunnel.

Multi-layer Label Stack

If a packet is transmitted in more than one layer of LSP tunnel, it will carry multiple layers of labels – Label Stack. At the ingress and egress of each tunnel, MPLS will PUSH or POP a label accordingly.

The label stack follows the "Last-In-First-Out" principle, so MPLS will process labels from the stack top.

MPLS sets no limit to the label stack depth. If the label stack depth of a packet is m, the label at the stack bottom is level 1, and the one at the stack top will be level m. A packet without pushing any label will be treated as having an empty label stack (the label stack depth is 0).

10.1.1.2 MPLS Network Introduction

MPSL Network Structure

As demonstrated in the next figure, the basic unit composing the MPLS network is LSR; and a

network consists of LSR is called a MPLS domain.

The LSR at the edge of a MPLS domain, connecting other customer networks is called LER (LER，Label Edge Router) , and the internal LSR is a core LSR. Core LSRs can either be routers supporting MPLS or ATM-LSR upgraded from ATM routers. LSRs in the domain communicate with each other via MPLS, while the MPLS domain edge is adapted via LER and traditional IP technologies.

Packets will be transmitted along a LSP composed of a series of LSRs after the ingress LER pushes a label to it. The ingress LER is called Ingress, egress LER called Egress, and routers in the middle called Trasit.

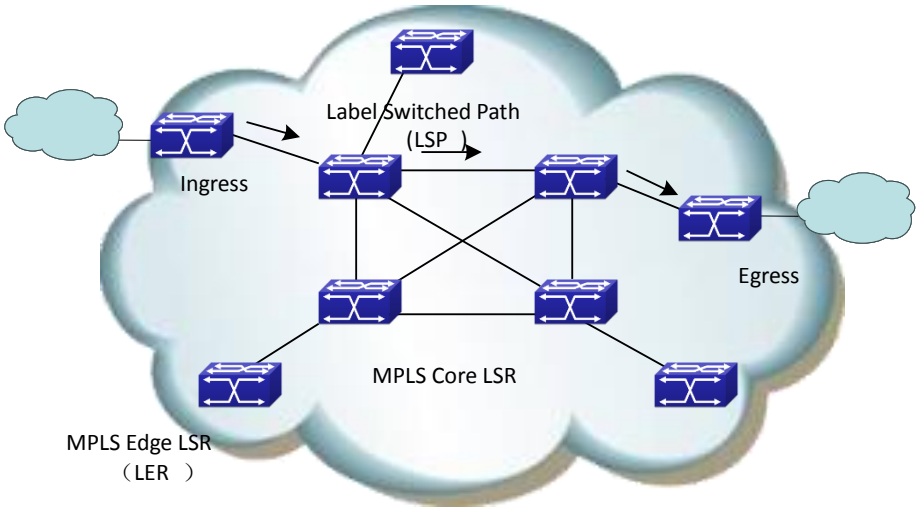


Figure 10-3 The MPLS Network Structure

The basic working process of MPLS based on the above figure :

First, LDP, together with traditional routing protocols (like OSPF, ISIS, etc) create route tables and LIB (Label Information Base) for FEC demanding services;

The ingress LER receives packets, completes L3 function, determines the FEC of the packets, labels them, and thus generates MPLS label packets.

Then, LSR in the network will forward packets according to their labels and LFIB (Label Forwarding Information Base) without implementing any L3 processing.

Finally, the egress LER of the MPLS will remove the label from the packet before the following IP forwarding.

To sum up, MPLS is neither a service or an application, but a tunnel technology, and a routing and switching technology platform integrated with label switching forwarding and network layer routing technology. This platform can support various high-level protocols and services with a certain guarantee of information security in the transmission.

10.1.1.3 Introduction to MPLS and Routing Protocols

When LDP creates LSP in hop-by-hop mode, it determines the next-hop based on the information from the forwarding table of each LSR along it. Since the information from forwarding tables are collected by routing protocols like IGP and BGP, LDP indirectly relates with them.

Besides, existing protocols like BGP and RSVP, can also distribute MPLS labels after extension.

Sometimes, it is necessary to extend some routing protocols in MPLS applications. For example, MPLS-based VPN requires extension to BGP, so that, BGP can distribute the VPN (Virtual Private Network) route information; MPLS-based TE (Traffic Engineering) requires extension to OSPF or IS-IS protocol, to carry link status information.

10.1.1.4 MPLS Application Introduction

MPLS technology originally combines L2 switching and L3 routing technology to enhance the route lookup speed. As ASIC (Application-Specific Integrated Circuit) develops, route lookup speed has no longer been the bottleneck of network development. As a result, MPLS's advantage in accelerating forwarding disappears.

However, combining the powerful L3 switching function of IP networks and efficient forwarding mechanism of traditional L2 networks, MPLS uses connection-oriented method at the forwarding plane, similar to the current L2 network. As a result, it can easily achieve seamless convergence of IP and L2 networks like ATM and Frame Relay, and provide better solutions for applications like QoS, TE and VPN.

MPLS-based VPN

Traditional VPN transmits private data in the public network via tunnel protocols like GRE, L2TP, and PPTP. Since LSP is a public network tunnel itself, MPLS is innately advantageous in implementing VPN.

MPLS-based VPN will connect different branches of a private network via LSP to form an integrated one. It also supports the intercommunication control between different VPN.

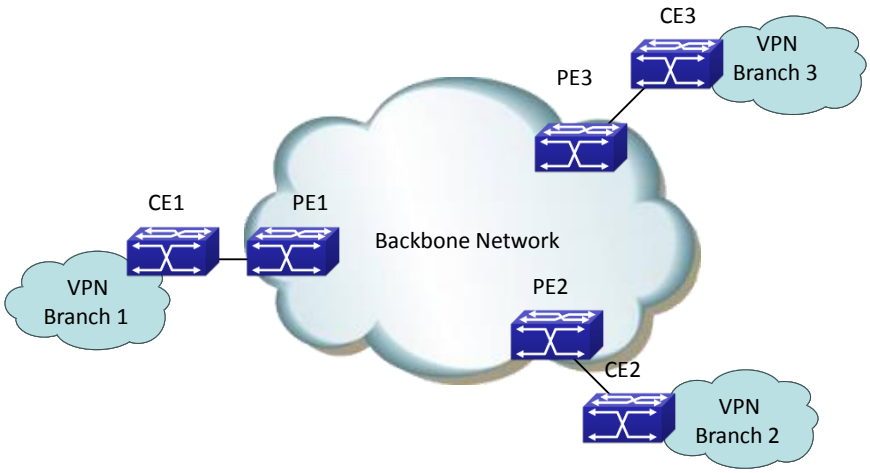


Figure 10-4 MPLS-based VPN

The above figure demonstrates the basic structure of MPLS-based VPN: CE (Customer Edge), a router, a switch or a host; PE (Provider Edge), in the backbone network.

PE manages VPN customers, establishes LSP connections between different PE and distributes routes to different branches of a VPN customer. The route distribution between PE is usually achieved via LDP or extended BGP.

MPLS-based VPN supports IP address multiplexing of different branches, and the

intercommunication between different VPN. Different with traditional routes, VPN route contains extra identification of branches and VPN, making BGP extension a necessity, in order to carry VPN route information.

MPLS-based TE

MPLS-based TE and the Diff-serv feature can provide data flow at different precedent level with different service while ensuring a high network utility efficiency, and hence, be able to provide low-delay, low packet loss rate services with a guaranteed bandwidth to various data flows like voice and video.

Considering the difficulty of deploying TE over the whole network, the Diff-serv model is usually the method of implementing QoS in real networking resolutions.

The basic mechanism of Diff-Serv is mapping a service to a certain service class at the network edge, according to the required service quality. The service is uniquely identified via the DS segment (originated from ToS field) of IP packet. According to the segment, the routers in the backbone network will apply pre-configured service policy to different services, ensuring the service quality.

The service quality class mechanism and the label mechanism of Diff-Serv are similar to the label distribution mechanism of MPLS. In fact, the MPLS-based Diff-Serv is implemented via the combination of the DS distribution and MPLS label distribution.

10.1.1.5 MPLS PHP

In the MPLS network, the core LSR will forward packets according to their labels. The Egress router (Egress LER) will remove the label before implementing IP forwarding.

In fact, in simple MPLS applications, where the Egress routers only implement IP forwarding, labels will become useless. In such cases, popping the labels out via the Penultimate Hop Popping feature at the penultimate router will stop the Egress router from processing the labels.

10.2 LDP

10.2.1 LDP Introduction

LDP protocol is used for label distribution in the MPLS label switching environment, and only applies to networks capable of label switching. LDP, integrated with traditional routing algorithm, distribute labels, advertise <label, FEC> map, create and maintain Label Forwarding Information Base and LSP, by transmitting various messages via TCP connections. LDP is used to distributing public network label in the MPLS VPN environment.

LDP doesn’t create any route; instead, it obtains routes from the system, distributes labels for them and advertises the labels to its upstream router. At the same time, for the FEC having a downstream, LDP will receive a label from the downstream, take it as the outgoing label and create a label switched path, which means to create an entry of

switching the incoming label as the outgoing one. If the label distributed by the downstream is 3, the LDP will create an entry of popping out the label.

LDP is defined in RFC3036; and its latest standard is RFC5036. It switches the map between labels and routes via the TCP connection between peers. Two neighbor discovery modes are supported by LDP: the basic mode (automatic discovery) and the extended mode (specified). The automatic discovery of peers is implemented via the UDP multicast messages to all routers (224.0.0.2), using the port 646 in both TCP and UDP messages.

The main process is as follows:

- ☞ Discover and maintain neighbors: after LDP is global enabled and interface enabled, it will send multicast Hello messages on the specified interface (unless it disables the multicast-based neighbor discovery) to advertise the network about its existence. The Hello messages will carry its transmission address, the address for TCP connections. The adjacency will be created when receiving Hello messages from other LSRs, and maintained by periodically sending Hello messages.
- ☞ Establish and maintain sessions: LDP sessions are TCP-based; First, compare the transmission address in the Hello message from the other end and that from this end, set the one with bigger value as ACTIVE and the other PASSIVE. The ACTIVE router will initiate a connect request to establish a TCP connection (to avoid the similar connection conflict problems suffered by BGP neighbors). Once the TCP connection is established, the two parts will send initialization messages to negotiate session parameters. A session will be established once the negotiation succeeds. After that, the two neighbors will send the local interface address list and label information to each other. To hold the connection when there is no data, KEEPALIVE messages will be sent.
- ☞ Create and maintain LSP: a session is necessary for each pair of LSR peers to switch label information, which create LSP by switching FEC and label binding messages.
- ☞ Cancel sessions: Without any message from the other end for a long time, LDP will disconnect the session and notify the close of the session to the other end by sending a notification messages.

Please notice that, LDP won't distribute labels for default routes, or BGP routes (unless explicitly specified).

10.2.1.1 Basic Concept of LDP

LDP Peer

When distributing labels to FEC, LDP needs to advertise this label and its meaning in the MPLS network to create LSP. LSR is a LDP peer when switching label information via LDP. LDP peers obtain each other's label map and other messages.

LDP Session

Two LSR will create a LDP session between each other after exchanging LDP Discovery Hello messages. LSP relies on LDP sessions to exchange messages like label map, release.

Two steps to establish a LDP session

- ☞ Establish the transmission connection.
- ☞ Initialize the session

Two types of LDP session:

- ☞ Local LDP Session: the two LSR establishing the session are directly connected.
- ☞ Remote LDP Session: the two LSRs establishing the session are indirectly connected

LDP Message Type

Four types of LDP messages:

- ☞ Discover message: to advertise and maintain the existence of LSR in the network;
- ☞ Session message: to create, maintain and terminate the sessions between LDP peers;
- ☞ Advertisement message: to create, change and delete the map from label to FEC;
- ☞ Notification message: to provide advice messages and error notices.

To ensure the reliable sending of LDP messages, LDP uses TCP to send Session, Advertisement and Notification messages, and UDP to send Discovery ones.

10.2.1.2 Introduction to LDP Message Format

LDP PDU

LDP PDU includes a LDP header and several LDP messages. The LDP header format is as follows:

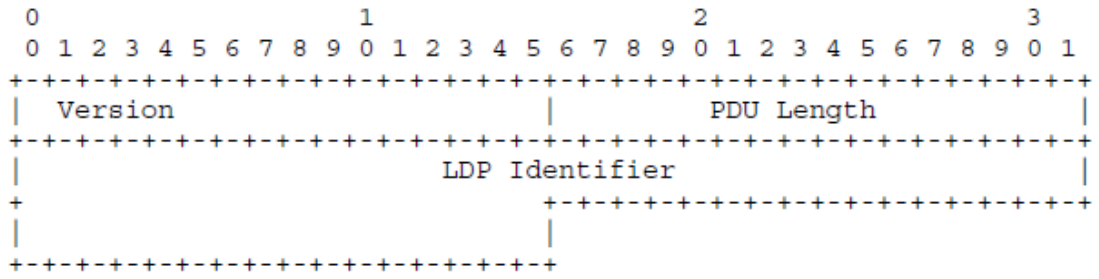


Figure 10-5 The LDP Header Format

- ☞ Version：The LDP version, I byte. The current LDP version is 1.
- ☞ PDU Length：The total length of the LDP message (in byte), 2 bytes.
- ☞ LDP ID：LDP ID, 6 bytes. The first 4 bytes is the globally unique LSR ID, and the rest 2 are label space ID, which is 0 when it comes to the global label space.

TLV Encoding

LDP encapsulates parameters in LDP messages via TLV (Type-Length-Value). The LDP TLV format is as follows:

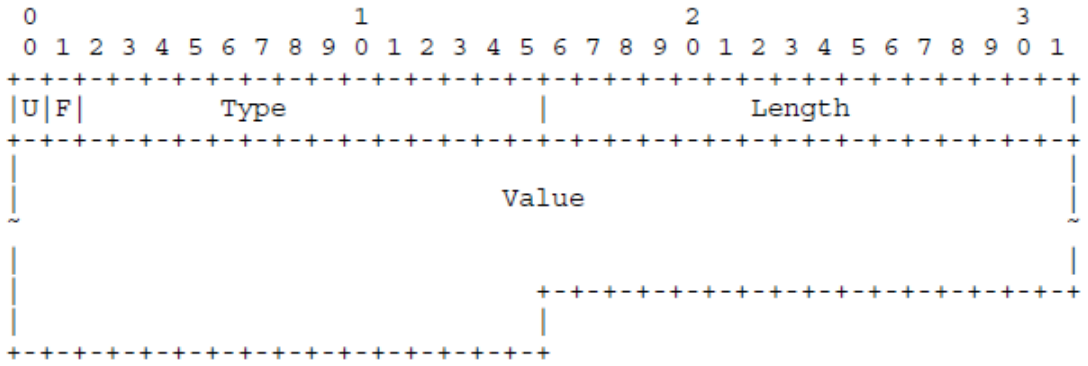


Figure 10-6 The TVL Format of LDP

- U bit：Unknown flag, I bit. If the U flag is 0, LSR should notify the source LSR of the packet and ignore the whole message; otherwise, ignore this TLV parameter and analyze other ones normally.
- F bit：Forwarding unknown TLV flag, 1bit. This flag only applies to LDP messages with unknown TLV and a U bit set as 1. If the F flag is 0, stop forwarding unknown TLV parameters; otherwise, forward them;
- Type：Type, 14 bits.
- Length：Length, 1 byte. The length of TLV value segment.
- Value：The Value segment, whose length is defined by the parameter of “Length”.
- The Value segment of TLV can also contain TLV parameters, meaning that, TLV are embeddable. The first byte of TLV doesn’t need alignment.

Currently defined TLV types:

TLV	Type
FEC	0x0100
Address List	0x0101
Hop Count	0x0103
Path Vector	0x0104
Generic Label	0x0200
ATM Label	0x0201
Frame Relay Label	0x0202
Status	0x0300
Extended Status	0x0301
Returned PDU	0x0302
Returned Message	0x0303
Common Hello Parameters	0x0400
IPv4 Transport Address	0x0401
Configuration Sequence Number	0x0402
IPv6 Transport Address	0x0403
Common Session Parameters	0x0500

ATM Session Parameters	0x0501
Frame Relay Session Parameters	0x0502
Label Request Message ID	0x0600
Vendor-Private	0x3E00- 0x3EFF
Experimental	0x3F00- 0x3FFF

10.2.1.3 LDP Label Management

In the MPLS system, the downstream LSR determines the distribution of label to specific FEC, and notifies the upstream. That is to say the labels are specified by the downstream and distributed from downstream to upstream.

Label Advertisement Mode

In the MPLS domain, packets will be forwarded to the downstream LSR with the downstream LSR label after the label switching process in the upstream LSR. The FEC labels distributed by the downstream LSR apply only to itself and the upstream LSR, and should be advertised to the upstream LSR. MPLS defines two label advertisement modes for the downstream LSR passing labels to its upstream LSR:

- DoD（Downstream On Demand）: LSR only distributes and advertises a label for the specified FEC after receiving a label request message from the upstream.
- DU（Downstream Unsolicited）: LSR distributes and advertises a label for the specified FEC without receiving a label request message from the upstream. It will automatically send label map information and notify the upstream LSR.

These two modes can be mixed, with each LSR interface configured independently to use one of them. During initialization, the upstream and downstream LSR have to exchange their label advertisement mode information to reach an agreement on the mode. Otherwise the creation of LSP will fail.

The figure demonstrates the process of LDP label advertisement:

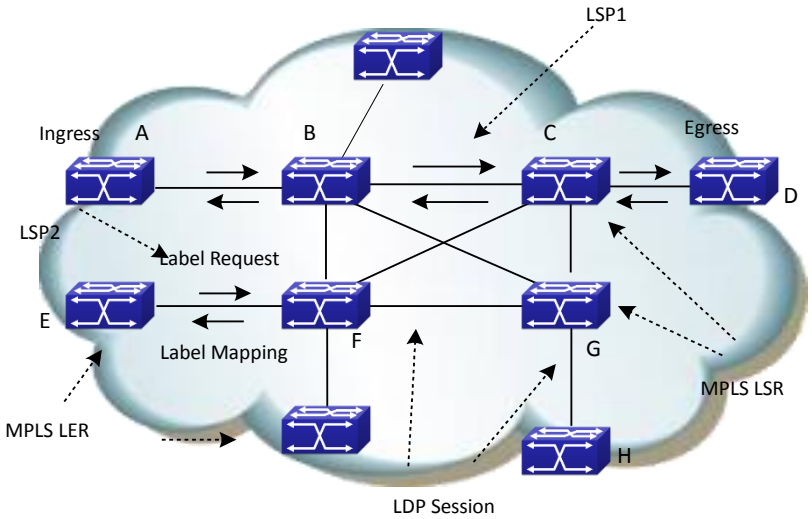


Figure 10-7 The Process of Label Advertisement

For example, as for LSP1 in the above figure, LSR B is the upstream LSR of LSR C, while LSR C is the downstream LSR of LSR B.

The main difference of two label advertisement mode lies on whether the label advertisement is DoD or DU.

The following is the detailed label advertisement process of these two modes:

DoD（downstream-on-demand）

The upstream LSR send a Label Request Message, which carries FEC description to its downstream LSR. The downstream LSR will distribute a label for this FEC, and respond to the upstream with the mapped label via a Label Mapping Message.

When will the downstream LSR respond the Label Mapping Message depends on the label advertisement mode it adopted.

In Ordered mode, it will send the Label Mapping Message to the upstream only after receiving a Label Mapping Message from its downstream.

In Independent mode, it will immediately send the Label Mapping Message to the upstream no matter it receives a Label Mapping Message from its downstream or not.

Usually, the upstream LSR chooses the downstream LSR based on the routing table. In the above figure, all LSR along the LSP1 work in the ordered mode, while LSR F in LSP2 in the Independent mode.

DU（downstream unsolicited）

The downstream LSR will automatically advertise the label mapping message to its upstream LSR after the LDP session successfully created. The upstream LSR will save the message and process it according to its retention mode.

Label Distribution Control Mode

In the MPLS domain, LSR generate a LSP from the ingress router and the egress router via switching labels, based on the route-forwarding path created by IGP in the MPLS domain. Only a complete path is useful for pack forwarding. The creation of LSP is the LSR label advertisement process; hence, controlling the creation of LSP is controlling the LSR label advertisement. MPLS defines two LSP control modes to determine when the downstream LSR will advertise labels to the upstream LSR:

- Ordered Mode: For a FEC label mapping of a LSR, the LSR only advertise the mapping to its upstream when it already has the label mapping of the FEC next-hop, or when it is the egress router of the FEC. The label advertisement of a flow starts from the egress router of this FEC flow, binding routers from downstream to upstream, thus to guarantee the mapping between labels and the flow is complete and coherent in the whole network. The ordered mode can prevent loop more effectively.
- Independent Mode: LSR doesn't have to wait for the label of the FEC next-hop to advertise labels to its peer. It can notify label mapping to the LSR connected to it at any time. This mode may cause the LSR advertise a label to its upstream before receiving one from its downstream. This mode can accelerate the creation

and aggregation of LSP.

Requirements for LSR to be an Egress router:

- ☞ The FEC quotes the LSR address;
- ☞ The FEC next-hop router locates outside the label switching network;
- ☞ The FEC unit passes the route area, such as another OSPF SUMMAERY domain, or another autonomy system of OSPF, BGP.

Label Retention Mode

Label Retention Mode determines how the LSR handles the currently useless mapping from label to FEC it received. In DU mode, the upstream LSR may receive a large number of <FEC, label> map sets from the downstream LSR, in which case, only when the FEC in the map set is the local FEC next-hop of the upstream LSR, this map set is meaningful for the label forwarding. MPLS defines two label retention modes to determine the processing of currently useless map set.

- ☞ Conservative Mode: the LSR will reserve the label mapping received from the neighbor LSR no matter the neighbor is its next-hop or not. The advantage of this mode is that it only creates and maintain the labels that meaningful for data forwarding, a very significant feature when the label space is limited (ATM switching).
- ☞ Liberal Mode: the LSR only save label maps from the neighbor LSR which is its next-hop. The advantage of this mode is that the expense of processing route changes is very low; and the disadvantage is many useless labels will be advertised and maintained.

In the Liberal label retention mode, LSR can adapt rapidly to route changes; in the Conservative mode, LSR can distribute and save relatively less labels. The Conservative retention mode, together with the DoD mode, usually applies to LSR with limited label space.

Some Basic Concepts of Label Switching

- ☞ NHLFE: Next Hop Label Forwarding Entry. It is used to describe the operation to the label, including Push and Swap.
- ☞ FTN (FEC to NHLFE map): the process of mapping FEC to NHLFE on the Ingress router.
- ☞ ILM (Incoming Label Map): the process of mapping received labels to NHLFE by LSR.

The Label Switching Process

The Ingress LER divides the packets entering the network into FECs. The packets belonging to the same FEC will follow the same path - LSP, in the MPLS domain. LSR will distribute a label for the incoming FEC packet and forward it through the corresponding interface.

The detailed process of label switch is as follows:

- ☞ All LSRs along the LSP will create an ILM first, the entries in which are the rule of mapping the incoming labels.

- ☞ LSR will map the labels of received packets to NHLFE;
- ☞ LSR will find the corresponding NHLFE in the LIB based on the label, replace it with the new label and then forward the label packet.

10.2.1.4 LDP Session

There are four steps to establish a LDP session:

- ☞ Discover
- ☞ Establish and maintain the session
- ☞ Create LSP
- ☞ Cancel the session

Discover

At this step, the LSR will send Hello messages periodically to adjacent LSRs, notifying them about its existence, in order to establish a session. In the basic discover mechanism, LSR will discover its LDP peers automatically via this process without manual configuration. There are two discover mechanisms:

- ☞ Basic Discover Mechanism

The Basic Discover Mechanism is used to discover local LDP peers – LSRs directly connected via the link layer, and create a local LDP session. In this mode, the LSR will send LDP Link Hello messages periodically via UDP messages to the multicast address marked as “all routers in the subnet”.

LDP Link Hello messages carry the LDP ID of the interface and other related information. If the LSR receives a LDP Hello Message at an interface, it means that there is a LDP peer at this interface (Link Layer).

- ☞ Extended Discover Mechanism

The extended discover mechanism is used to discover remote LDP peers – LSRs not directly connected via the link layer, and created remote LDP sessions. In this mode, the LSR will send LDP Targeted Hello messages periodically to the specified IP address via UDP messages.

LDP Targeted Hello messages carry the LDP ID of the interface and other related information. If the LSR receives a LDP Targeted Message at an interface, it means that there is a LDP peer at Network Layer.

Establish and Maintain the Session

After discovering a LDP peer, LSR will began to establish the session in two steps:

- ☞ Establish the transmission layer connection, that is, a TCP connection between LSRs;
- ☞ Initiate the session between the LSRs, negotiate all concerning parameters, such as the LDP version, the label advertisement mode, the timer value, the label space. After the negotiation succeeds, the session is established between the LSRs.

The session will be maintained by Keepalive messages after established.

Create LSP

The process of creating LSP is mapping FEC and labels and advertising the maps to the adjacent LSRs along the LSP, which is realized via LDP. Take DoD mode as the example, the main steps are as follows:

When the network routes change, if an edge router finds out a new destination address in its route table which belongs to none of the existing FECs, it needs to create a new FEC for this destination address. The edge LSR determines the route for the FEC, initiates a label request message to its downstream LSR, and specifies for which FEC this label request is.

The downstream receiving the label request message will save this message, finds the corresponding FEC next-hop according to the local route table and then sends a label request message to its downstream.

When the label request message reaches the destination router or the egress router of the MPLS network, if the router has available labels, and judges the label request messages as legal, it will distribute a label for the FEC, and send a label mapping message containing the label information to its upstream;

The LSR receiving the label mapping message will check the state of label request messages saved locally. If there is a corresponding label request message of a FEC label mapping message in the data base, LSR will distribute a label for the FEC, and add a new entry in its LFIB, and then send the label mapping information to its upstream.

When the ingress LSR receives a label mapping message, it also should add a corresponding entry in its LFIB, and thus finish the creation of LSP.

Cancel the session

LDP maintains adjacency by checking Hello messages. It also maintains session by checking Keepalive messages. If there is no Keepalive message received within a certain period of time, the LDP session will close the connection.

Each LDP session can include one or more Hello adjacencies. LDP maintains Hello adjacency via periodical Hello messages. If there is no LDP Discovery Hello message received within a certain period of time, the LDP session will close the Hello connection. When closing the last Hello adjacency in the LDP session, LDP will send notification messages, and close the transmission connection.

10.2.1.5 LDP Loop Detection

Creating LSP in the MPLS domain also needs to prevent loops. The LDP loop detection mechanism can detect LSP loops and avoid them.

To detect loops in the MPLS domain, all LSRs should be enabling the loop detection. But when establishing LDP sessions, the configurations of loop detection on the two parties don’t have to be the same.

There are two LDP loop detection modes:

The maximum hop count

It is the number of LSR passed by the label messages (including label mapping and

label request). When LSR transmits label information with the hop-count parameter, it will first increase the hop count by 1. When the hop count reaches the configured maximum value, it means that a loop exists, and the LSP creation will fail. If the hop count is 0, it means the hop count is unknown. The hop count of label messages is always 0. The default maximum hop count is 255.

Path Vector

It is used to record the path information in label mapping or label request messages. At each hop, the LSP checks whether its LSR ID is in the record. The following two conditions mean the existence of a loop and the failure of the LSP creation.

- ☞ There is a record of this LSR in the path vector record;
- ☞ The hop count of the path exceeds the configured maximum value.

If no record of its LSR ID is found, a new one will be added. The maximum value of path vector is the same as that of the hop count.

10.2.2 LDP Configuration

LDP Configuration Task Sequence:

Enable MPLS Globally (Necessary)

Enable LDP (Necessary)

Enable/Disable the LDP module

Enable/Disable label-switching on the interface

Enable/Disable LDP module on the interface

Configure the LDP parameters (optional)

Configure the LDP label management mode

Configure the LDP label retention mode

Configure the LDP label advertisement mode

Configure the LDP label control mode

Configure the LDP loop detection

Enable/Disable the LDP loop detection

Set the maximum hop count of the LDP loop detection

Configure the LDP specified peers

Configure other LDP parameters

Configure the aging time or interval of each timer

ID Configure the LDP router ID

Configure the TCP interface address of LDP

Configure the LDP to discover peers via multicast Hellos or not.

Configure the LDP to import BGP routes or not.

Enable/Disable the LDP label merging capability

Configure the LDP to transmit release messages or not.

Configure the LDP to retry or not when the label request is rejected.

Hello Configure the LDP to receive Hello from specified targets

Clear LDP connections or adjacencies.

Globally enable MPLS

Command	Explanation
Global Mode	
mpls enable **no mpls enable**	necessary Enable MPLS; the no operation will disable MPLS.

Enable LDP

It is easy to implement basic configurations of LDP in SNR. Usually users only have to enable the LDP switch, and enable it on the interface where the LDP will work. Please notice that, the interface with LDP enabled should enable label switching.

Command	Explanation
Global Mode	
router ldp **no router ldp**	Necessary LDP Enable/disable LDP; disabled by default
Interface Configuration Mode	
label-switching **no label-switching**	Necessary Enable/disable label-switching; disabled by default
mpls proxy loopback-group <1-max_agg_num> **no mpls proxy loopback-group**	Enable MPLS proxy, the default does not enable the function, enable MPLS proxy when the boardcard of MPLS ingress unicast packet does not support MPLS.
ldp {enable|disable}	Necessary LDP Enable/disable LDP on the interface; disabled by default

Configure the LDP parameters

Configure the LDP label management mode

Configure the LDP label retention mode

Configure the LDP label advertisement mode

Configure the LDP label control mode

Command	Explanation
Router Configuration Mode	
label-retention-mode {conservative|liberal}	Optional Configure the global label retention mode: Conservative or Liberal; it is liberal by default

advertisement-mode {downstream-on-demand|downstream-unsolicited}	Optional Configure the global label advertisement mode: **downstream-on-demand** or **downstream-Unsolicited.** This mode relates with the other two. The change of it will change the label retention mode and the global label path control mode at the same time. It is **downstream-unsolicited** by default
control-mode {ordered|independent}	Optional Configure the global label retention mode: Ordered or **independent**; it is **independent** by default
Interface Configuration Mode	
ldp label-retention-mode {conservative | liberal}	Optional Configure the label retention mode of the interface; the default value is the same as the global configuration. If the configuration differs with the global one, the interface configuration will take effect.
ldp advertisement-mode {downstream-on-demand|downstream-unsolicited}	Optional Configure the label advertisement mode of the interface; the default value is the same as the global configuration. If the configuration differs with the global one, the interface configuration will take effect.

Configure LDP loop detection

Enable/disable LDP loop detection

Configure the maximum hop count of LDP loop detection

Command	Explanation
Router Configuration Mode	
[no] loop-detection	Optional Enable LDP loop detection, the no operation will disable it.
[no] loop-detection-count <*count*>	optional Configure the maximum hop count of LDP loop detection, whose default value is 255, the no operation will restore the default value.

Configure the LDP specified peers

Command	Explanation
Router Configuration Mode	
[no] targeted-peer <*ip-addr*>	optional Configure the remote peer of the LDP targeted destination.

Configure other LDP parameters

Configure the aging time or interval of each LDP timer

ID Configure LDP router ID

Configure the TCP interface address of LDP

Configure the LDP to discover peers via multicast Hellos or not,

Configure the LDP to import BGP routes or not.

Configure the LDP to enable label merging capability or not.

Configure the LDP to transmit release messages or not.

Configure the LDP to retry or not when the label request is rejected

Hello Configure the LDP to receive Hello from the specified targets

Command	Explanation
Route Configuration Mode	
[no] keepalive-interval <*interval*>	Optional Configure the interval of sending LDP keepalive messages, whose default value is 10 seconds; the no operation will restore the default value
[no] keepalive-timeout <*time-val*>	Optional Configure the LDP keepalive timeout, whose default value is 30 seconds; the no operation will restore the default value
[no] Hello-interval <*Hello-interval*>	Optional Configure the interval of sending multicast HELLO messages, whose default value is 5 seconds; the no operation will restore the default value
[no] hold-time <*hold-time* >	Optional Configure the LDP multicast peer hold time, whose default value is 15 seconds; the no operation will restore the default value
[no] targeted-peer-Hello-interval <*Hello -interval*>	optional Configure the interval of sending HELLO to specified targets, whose default value is 15 seconds; the no operation will restore the default value
[no] targeted-peer-hold-time <*hold-time*>	optional Configure the LDP targeted peer hold time, whose default value is 45 seconds; the no operation will restore the default value
Interface Configuration Mode	
[no] ldp keepalive-interval <*interval*>	optional Configure the interval of sending LDP keepalive messages on a specified interface; the no operation will restore the default value

[no] ldp keepalive-timeout <*time-val*>	Optional Configure the LDP keepalive timeout on a specified interface; the no operation will restore the default value
[no] ldp Hello-interval <*Hello-interval*>	Optional Configure the interval of sending LDP multicast HELLO messages on a specified interface; the no operation will restore the default value
[no] ldp hold-time <*hold-time*>	optional Configure the LDP multicast peer hold time on a specified interface; the no operation will restore the default value
[no] ldp targeted-peer-Hello-interval <*Hello-interval*>	optional Configure the interval of sending LDP HELLO messages to specified targets on a specified interface; the no operation will restore the default value
[no] ldp targeted-peer-hold-time <*hold-time*>	optional Configure the LDP targeted peer hold time on a specified interface; the no operation will restore the default value
router configuration mode	
[no] router-id <*ip-addr*>	optional Configure the LDP router ID, which is obtained automatically by default. The no operation will cancel the manually configured router ID, and automatically obtain a valid interface IP address as the router ID.
[no] transport-address <*ip-addr*>	optional Configure the IP address of LDP for TCP connections. Please notice that this address has to be that of a loopback interface on the main VRF. The no operation will cancel the manual configuration and let LDP automatically choose the TCP address
[no] multicast-Hellos	optional Configure the LDP to discover peers via multicast HELLOs, the no operation will do the opposite. Using multicast HELLO is the default setting.
[no] import-bgp-routes	Optional Configure the LDP to import BGP routes; the no operation will do the opposite. Not importing BGP routes is the default setting.

[no] global-merge-capability {merge-capable|non-merge-capable}	optional Configure the LDP to enable global label merging capability or not, the no operation will restore the default value.
[no] propagate-release	optional Configure the LDP to advertise label release messages to peers, the no operation will do the opposite. Not transmitting label release messages is the default setting.
[no] request-retry	optional Configure the LDP to retry 5 times when the label request is rejected, the no operation will disable the retry.
[no] request-retry-timeout <*time-val*>	optional Configure the retry interval, whose default value is 5 second, the no operation will restore the default value.
[no] targeted-peer-Hello-receipt	optional Configure LDP to receive HELLOs from specified targets, even the targeted peer is not configured on the host. Not receiving such HELLOs is the default setting. The no operation will restore the default configuration. Please notice that, if targeted LDP peers are configured, targeted-peer-Hello-receipt should be too.

Clear the LDP connections or adjacencies

Command	Explanation
Admin Mode	
clear ldp adjacency {<*ip-addr*>|*}	Optional Clear specified LDP adjacencies, “*” means all.
clear ldp session {<*ip-addr*>|*}	optional Clear specified LDP sessions, “*”means all.

10.2.3 LDP Typical Instances

Some designations of LDP are for adapting different network environments. Its configuration is very simple in the typical Ethernet environment. Due to the development of hardware system, especially the popularity of L3 switches, the pure MPLS network has already lost its importance to MPLS VPN.

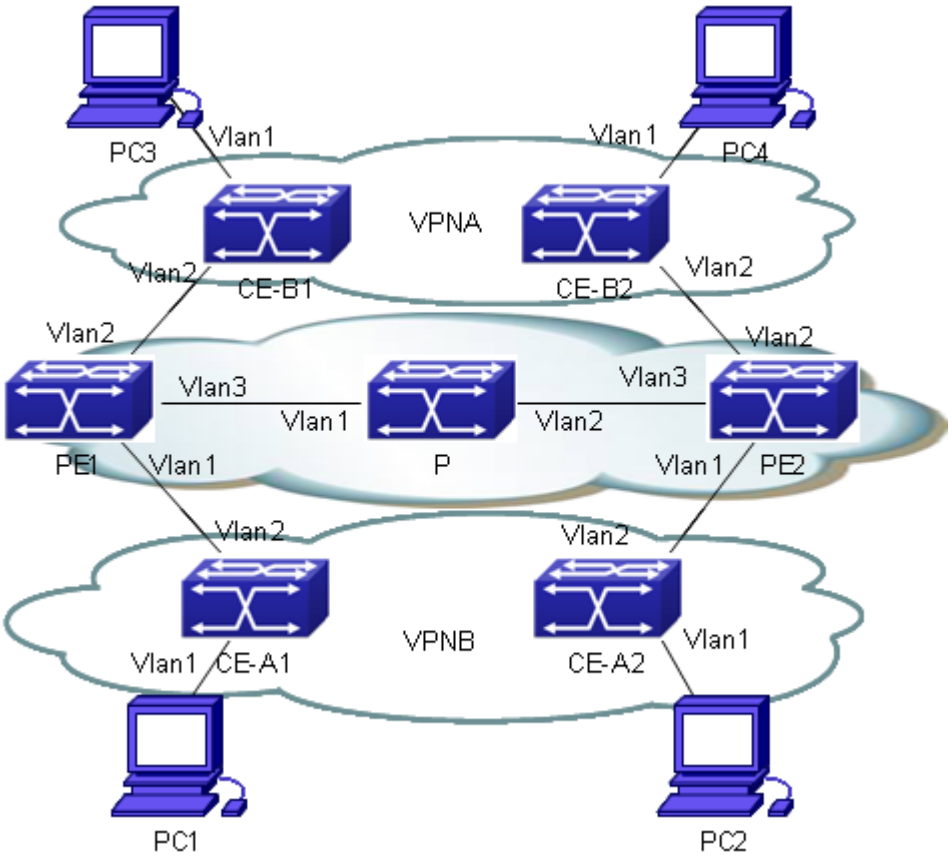


Figure 10-8 MPLS VPN Typical Instance

The above figure demonstrates a typical MPLS VPN instance, in which, PE1, P and PE2 form the public network area – the area switching via MPLS. CE-A1 and CE-A2 form VPN-A, CE-B1 and CE-B2 form VPN-B. Both VPNs communicate via the public network label switching, and need to configure LDP for distributing and advertising labels in the public network area. To guarantee the reachability of routes, we advertise routes via OSPF.

The LDP configuration of PE1 is as follows:

```
PE1#config
PE1(config)#mpls enable
PE1(config)# router ldp
PE1(config-router)#exit
PE1(config)#interface vlan 3
PE1(config-if-Vlan3)#ip address 202.200.1.2 255.255.255.0
PE1(config-if-Vlan3)#ldp enable
PE1(config-if-Vlan3)#label-switching
PE1(config-if-Vlan3)#exit
PE1(config)#router ospf
PE1(config-router)#network 200.200.1.1/32 area 0
PE1(config-router)#network 202.200.1.0/24 area 0
PE1(config-router)#exit
```

The LDP configuration of P is as follows:

```
P#config
```

P(config)#mpls enable
P(config)# router ldp
P(config-router)#exit
P(config)#interface vlan 1
P(config-if-Vlan1)#ip address 202.200.1.1 255.255.255.0
P(config-if-Vlan1)#ldp enable
P(config-if-Vlan1)#label-switching
P(config-if-Vlan1)#exit
P(config)#interface vlan 2
P(config-if-Vlan2)#ip address 202.200.2.1 255.255.255.0
P(config-if-Vlan2)#ldp enable
P(config-if-Vlan2)#label-switching
P(config-if-Vlan2)#exit
P(config)#router ospf
P(config-router)#network 202.200.1.0/24 area 0
P(config-router)#network 202.200.2.0/24 area 0
P(config-router)#exit

The LDP configuration of PE2 is as follows:

PE2#config
PE2(config)#mpls enable
PE2(config)# router ldp
PE2(config-router)#exit
PE2(config)#interface vlan 3
PE1(config-if-Vlan3)#ip address 202.200.2.2 255.255.255.0
PE2(config-if-Vlan3)#ldp enable
PE2(config-if-Vlan3)#label-switching
PE2(config-if-Vlan3)#exit
PE2(config)#router ospf
PE2(config-router)#network 200.200.1.2/32 area 0
PE2(config-router)#network 202.200.2.0/24 area 0
PE2(config-router)#exit
Please refer to BGP VPN typical instances for the configuration of BGP.

10.2.4 LDP Troubleshooting

When configuring and using LDP, some problems like incorrect physical connections, configuration errors may cause it to fail, so plesase pay attention to the following notices to avoid them:

- ☞ First, make sure the system enables LDP globally and on the active interface. Notice that the LDP can only be enabled on interfaces after it is enabled globally.
- ☞ Second, use the “show ldp interface” command to check whether the LDP has been enabled correctly on the interface after the connection succeeds. If the LDP has been correctly enabled but cannot be displayed, it is possible that the

interface is not in the UP mode or not configured with interface label-switching.

- ☞ Then, make sure the adjacent interfaces are in the same segment, and check whether the LDP can discover peers and establish adjacencies with them normally via the “show ldp adjacency” command. If no peer is discovered or no adjacency is established, it is possible that the interfaces may belong to different segments, or one of the local host and its remote neighbor disables multicast HELLO. Besides, when establishing TCP connection, LSR ID is the default address, as, please make sure advertise the LSR ID route to the remote end.
- ☞ Check whether the state of LDP session with “show ldp session” is operational, since only in this state, LDP sessions can switch messages. If the LDP session can’t be established, use “show ldp” to check the TCP addresses of the two parties, and lookup the route table to make sure the route of the remote end is reachable.

At last, given all above steps succeed, use “show ldp fec” to check the routes imported by LDP and their information, or check the created entries with “show mpls ftn” and “show mpls ilm”.

10.3 MPLS VPN

10.3.1 BGP/MPLS VPN Introduction

10.3.1.1 BGP/MPLS VPN Network Structure

BGP/MPLS VPN is a PE-based L3VPN technology in the VPN solutions provide by providers, using BGP to advertise VPN routes and MPLS to forward VPN messages in the provider backbone network.

The BGP/MPLS VPN networking is flexible, extendible, and can support MPLS QoS and MPLS TE conveniently, resulting in its increasingly popular application.

BGP/MPLS VPN model consists of three parts: CE, PE and P.

- ☞ P router: Provide Router. It locates in the MPLS domain, and is able to switch fast-forwarding MPLS data flow based on labels. P router receives MPLS messages, switch labels and then output them.
- ☞ PE router: Provide Edge Router. It locates at the edge of the MPLS domain, for converting IP messages and MPLS messages. PE router receives IP messages, pushes MPLSU labels, and output MPLS messages; or receives MPLS messages, pop labels, and output IP messages. On PE routers, the ports connected with other P routers or PE routers are “public network port”, configured with public network IP address; those connected with CE routers are “private network port”, configured with private network address.

- CE router: Customer Edge Router. It locates at the edge of the customer IP domain, connected directly to PE route, for aggregating customer data and forwarding route information of the customer IP domain to PE router.

The next figure demonstrates a BGP/MPLS VPN networking:

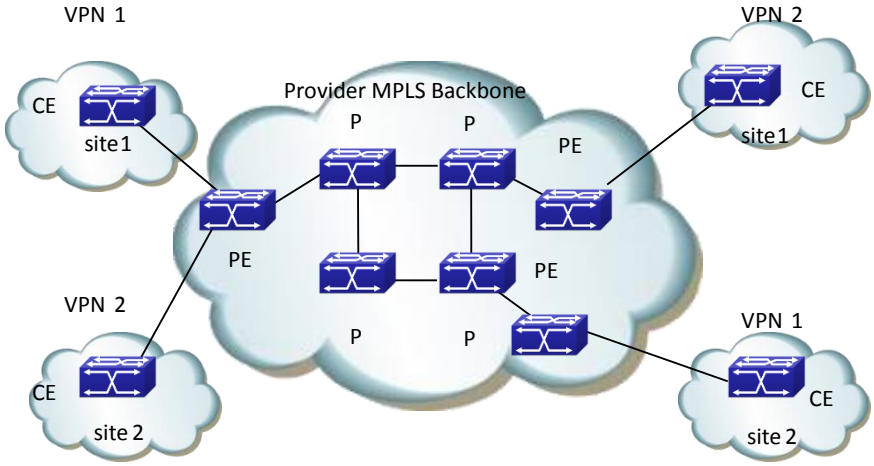


Figure 10-9 BGP/MPLS VPN Networking

The division of CE and PE is based on the management areas of SP and customers, since CE and PE are the edge between the two areas.

CE is usually a router. When the adjacency between it and the PE directly connected to it, CE will advertise the local VPN route to PE, and learn the remote VPN route from PE. CE and PE use BGP/IGP to exchange route information or static routes.

PE will exchange VPN route information with other PEs via BGP after learning the local VPN route form CE. It only maintains the VPN route directly connected with it rather than all VPN routes in the service provider network.

P router only maintains routes to PE, without learning any VPN route information.

Then transmitting VPN traffic in the MPLS backbone network, the ingress PE serves as the Ingress LSR (Label Switch Router), the egress PE the Egress LSR, and P router the Transit LSR.

10.3.1.2 Basic Concept of BGP/MPLS VPN

Site

"Site" is a concept usually mentioned when introducing VPN, which can be understood from the following aspects:

- Site is a set of IP systems with IP connectivity between each other. This connectivity is independent of SP network.
- The division of site is based on the topology of devices instead of devices' location, although in most cases, the devices in a site locate next to each other.
- The devices in a site can belong to multiple VPN. In other words, a site can

belong to multiple VPN;

☞ Site connects to SP network via CE. One site can include multiple CE while a CE can only belong to one site.

Multiple sites connected to the same SP network can be divided into different sets according to special policies, which only allow intercommunication via the SP network to happen between the sites within the same set. Such sets are VPN.

VRF

VRF (VPN Routing & Forwarding Instance), consisting of VPN IP route table and VPN IP forwarding table (the forwarding table contains the MPLS encapsulation information), is the core entry of MPLS VPN packet forwarding. Each VPN has its own independent VRF. The VRF address spaces of different VPN can overlap with each other. A PE/P router in the MPLS VPN network usually contains multiple independent VRF.

Overlapping Address Space

VPN is a private network, which means each VPN manages its own address range independently. This range is called Address Space.

The address spaces of different VPN may partially overlap with each other. For example, if VPN1 and VPN2 both use the segment of 10.110.10.0/24, there would be Overlapping Address Space.

VPN instance

In the MPLS VPN, the route isolation between different VPN is implemented via VPN instance.

PE creates and maintains a special VPN instance for every site directly connected to it. VPN site contains the VPN membership and route rules of the corresponding site. If the customers of a site belong to more than one VPN, then its VPN instance will contain the information of all those VPN.

To guarantee the data independency and security of VPN, each VPN instance on PE has its own independent route table and IFIL (Label Forwarding Information Base).

To be specific, the information in VPN instances include: LFIB, IP route table, interfaces bound with VPN instance, and its management information (including RD, route filter policy, member interface list and etc).

VPN-IPv4 Address

The traditional BGP can't correctly handle the VPN routes with overlapping address spaces. Assume that VPN1 and VPN2 both use the segment of 10.110.10.0/24, and advertise separately a route reaching this segment, BGP will only choose one of the two routes, losing the one reaching the other VPN.

PE routers use MP-BPG to advertise VPN routes between each other and solve the above problem via VPN-IPv4 address family.

A VPN-IPv4 address consists of 12 bytes, including 8 bytes of RD (Route Distinguisher) and 4 bytes of IPv4 address prefix.

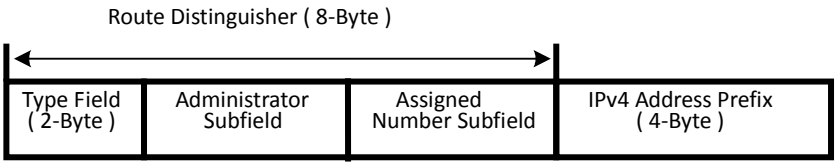


Figure 10-10 VPN-IPv4 Address Structure

After receiving the regular IPv4 routes from CE, PE should advertise these private network VPN routes to the remote PE. The independency of the private network routes is based on the additional RD patched to them.

SP can independently distribute globally unique RD, thus, even the VPN from different SP networks use the same IPv4 address space, the PE routers can advertise different routes to them.

It is recommended to allocate a special RD for each VPN instance on the PE to ensure all routes reaching the same CE uses the same RD. the VPN-IPv4 address whose RD is 0 is a globally unique IPv4 address.

Adding RD is to a specific IPv4 prefix will make the latter globally unique, which is the meaning of RD.

RD may relate with ASN, in which case, it is a combination of an ASN and a random number; it may also relate with IP address, in which case, it is a combination of an IP address and a random number.

There are two RD formats, differing with each other via 2 bytes of Type filed:

- If Type is 0, the Administrator sub-field takes up 2 bytes, Assigned Number sub-field takes up 4 bytes. The format would be: 16 bits of ASN: 32 bits of user-defined number. For example: 100:1
- If Type is 0, the Administrator sub-field takes up 2 bytes, Assigned Number sub-field takes up 4 bytes. The format would be: 32 bits of IPv4 address: 16 bits of user-defined number. For example: 172.1.1.1:1

To guarantee the global uniqueness of RD, please don't set the value of Administrator sub-filed as private ASN or private IP address.

VPN Target Attribute

BGP/MPLS VPN uses a 32 bit BGP extended community attribute – VPN Target (also called Route Target) to control the advertisement of VPN route information.

There are two types of VPN Target attribute used by VPN instances on PE routers:

- Export Target attribute: the local PE sets the Export Target attribute for the VPN-IPv4 routes it learns from the sites directly connected to it, before advertising the routes to other PE.
- Import Target Attribute: when receiving the VPN-IPv4 route advertised by other PE routers, PE will check their Export Target Attribute, and add the routes into corresponding VPN route table only when their Export Target attributes match the Import Target attributes of the VPN instances on it.

In other words, VPN Target attribute defines which sites can accept a VPN-IPv4 route, and a PE router can receive routes from witch sites.

Like RD, there are two VPN Target formats:

- ☞ 16 bits ASN : 32bits user-defined number, for example: 100:1
- ☞ 32bits IPv4 address: 16 bits user-defined number, for example: 172.1.1.1:1

MP-BGP

MP-BGP (Multiprotocol extensions for BGP-4) transmits VPN information and routes between PE routers. MP-BGP is backward-compatible, simultaneously supporting traditional IPv4 address family and other address family (such as VPN-IPv4 address family). It can ensure the advertisement of private network VPN routes only happens within the VPN, and can realize the communication between MPLS VPN members.

Routing Policy

On the basis of controlling VPN route advertisement via ingress and egress extended community, the import or export route policy can be used for a more precise control of importing and advertising VPN routes.

The import route policy can filter the routes importable for VPN instances according to the VPN target attribute of routes. It can deny the receipt of routes specified by the community in the import list. The export route policy can deny advertising the routes specified by the community in the export list.

After creating VNP instances, users can choose whether to configure import or export route policy.

Tunneling Policy

Tunneling Policy is used to choose tunnels for specified VPN instances messages.

Tunneling Policy is optional. After creating VNP instances, users can configure it. By default, it will choose LSP as the tunnel without load sharing (the load sharing number is 1). Besides, this policy only takes effect in one AS domain.

10.3.1.3 Forwarding BGP/MPLS VPN Messages

In basic L3VPN applications (not include Multi-AS VPN), the forwarding of VPN packets adopts the 2-layer label mode:

- ☞ The first layer (outer layer) labels will be switched within the backbone network, indicating a LSP from the PE to the remote PE. With this layer of label, VPN messages can reach the remote PE along the LSP.
- ☞ The second layer label (inner layer) will be used when the packet reaches CE from the remote PE, indicating which site to send the packet, or, more specifically, which CE it will reach. Thus the remote PE will find the correct interface to forward the packet according to this layer of label.

In some special conditions, two sites belonging to the same VPN may connect to a same PE, in which case, the only information matters is how to reach the remote CE.

The next figure demonstrates an example of forwarding VPN packets:

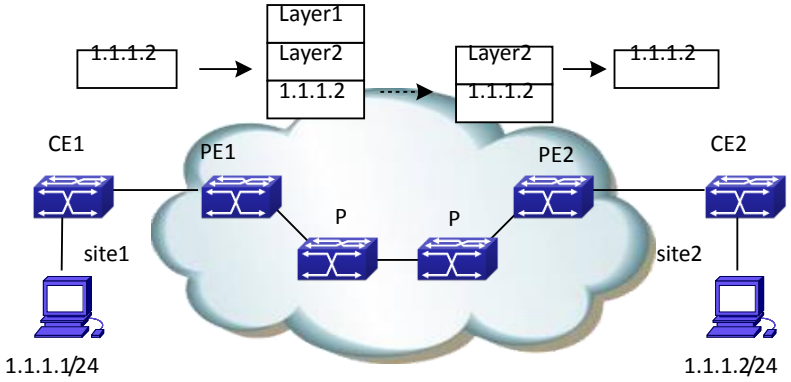


Figure 10-11 Forwarding VPN Packets

Site1 sends an IP packet with a destination address of 1.1.1.2, which is sent by CE1 to PE1.

PE1 looks up VPN-instance entries according to the interface receiving the packet and the destination address, then forwards the packet after adding two layers of label (inner and outer) to it, if there is a match.

The MPLS network will send the packet to PE2 according to the outer layer label (removed when the packet reaching the last-hop of PE2, leaving only the inner layer) of it.

PE1 looks up VPN-instance entries according to the inner layer of label and the destination address, then forwards the packet to CE2 after determining its egress interface.

CE2 forwards the packet to its destination according to the regular IP forwarding process.

10.3.1.4 BGP/MPLS VPN Networking Resolution

In BGP/MPLS VPN networks, the advertisement and receipt of VPN routes between different sites are controlled by VPN Target Attribute. The configurations of VPN Export Target and Import Target are independent, both allowing multiple values, and hence can realize flexible VPN access control and various VPN networking resolutions.

Basic VPN

In the most basic instance, all users of a VPN form a closed user group, allowing the forwarding of traffic between them. But no user within the VPN can communicate with outside users.

In such networking, each VPN will obtain an exclusive VPN Target as its Export Target and Import Target, which should not be used by other VPN.

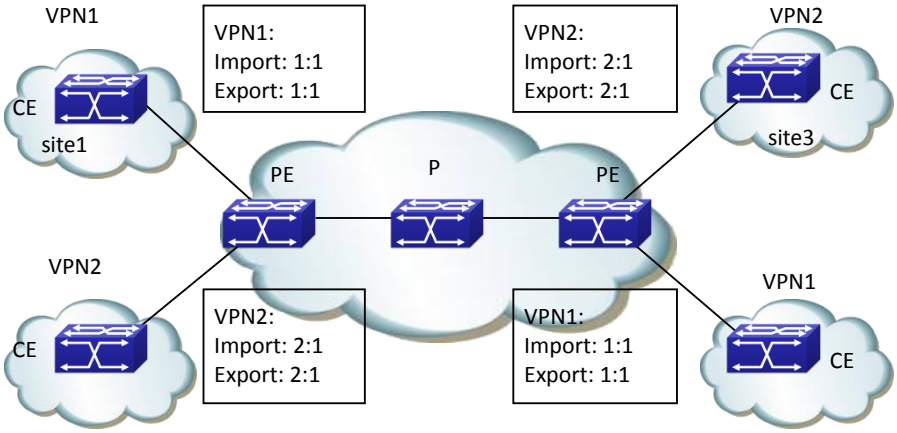


Figure 10-12 Basic VPN Networking Resolution

In the above figure, the VPN Target distributed by PE for VPN1 is 100:1; and that for VPN2 is 200:1. The sites of VPN1 can intercommunicate with each other, so do the two of VPN2. But the intercommunication between sites in VPN1 and those in VPN2 arise forbidden.

Hub&Spoke VPN

To use a central access control device in VPN to control the intercommunication of other users, Hub&Spoke networking resolution is a good choice, so that the central device can monitor and filter the intercommunication between the devices at two ends.

Two VPN target is needed in this networking, one for “Hub”, the other for “Spoke”.

All sites should follow the following rules to configure VPN Target for VPN instances on PE:

- Spoke-PE: Export Target is “Spoke”, Import Target is “Hub”
- Hub-PE: two interfaces or sub-interfaces are needed, one for receiving routes from Spoke-PE, the Import Target of whose VPN instance is “Spoke”; the other for advertising routes to Spoke-PE, the Export Target of whose VPN instance is “Hub”.

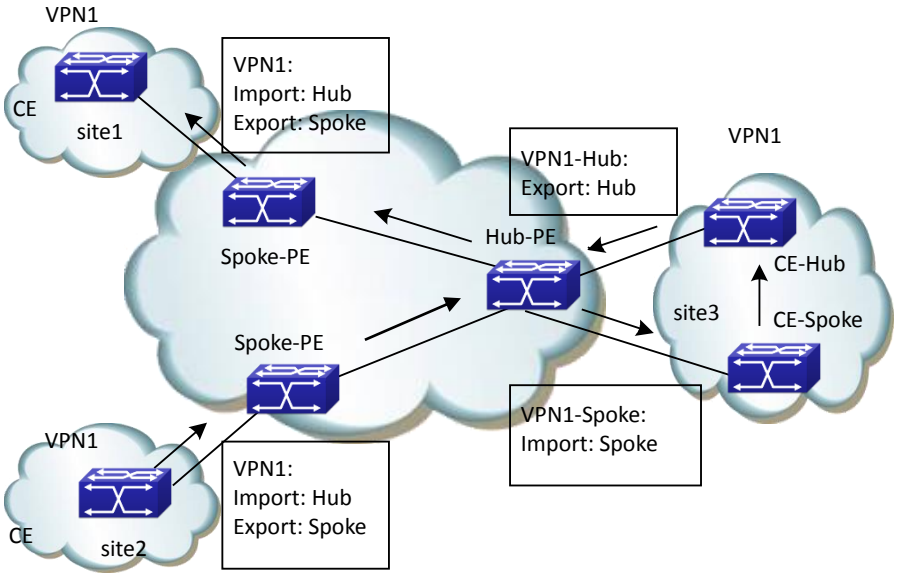


Figure 10-13 Hub&Spoke Networking Resolution

In the above figure, Spoke sites communicate with each other via Hub sites (the arrow in the figure is the route advertisement process from site2 to site1):

- Hub-PE can receive VPN-IPv4 routes advertised by all Spoke-PE
- The VPN-IPv4 routes advertised by Hub-PE can be received by all Spoke-PE;
- Since Hub-PE can advertise routes it learns from Spoke-PE to other Spoke-PE, the spoke sites can intercommunicate with each other via the Hub site.
- The Import Target attribute of any Spoke-PE is different from the Export Target attribute of other Spoke-PE. So, any pair of Spoke-PE cannot advertise VPN-IPv4 routes to each other or intercommunicate directly.

Extranet VPN

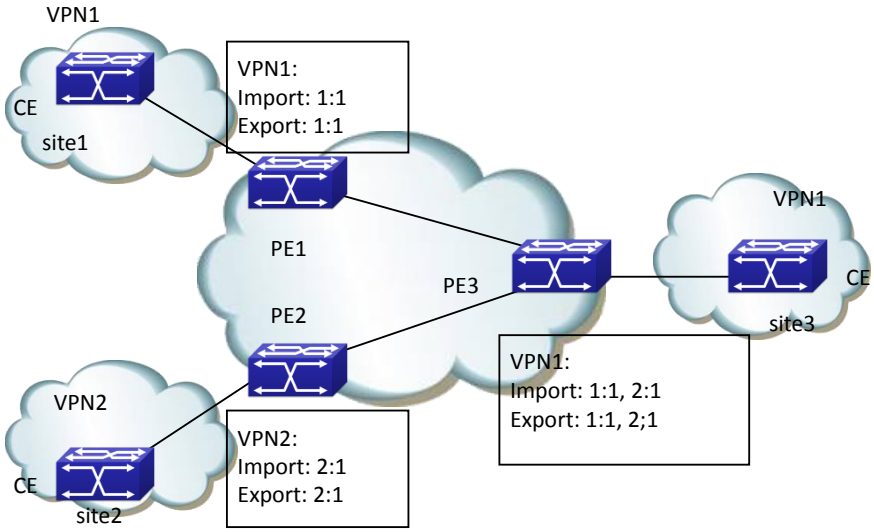


Figure 10-14 Extranet Networking Resolution

If a VPN user wants to provide some site resource of this VPN to outside users, the Extranet Networking resolution can solve the problem.

In this networking if a VPN needs to access the sharing site, its Export Target should be included in the Import Target of the sharing site VPN instances, and its Import Target should be included in the Export Target of the sharing site VPN instances.

In the above figure, site3 of VPN1 can be accessed by VPN1 and VPN2:

- PE3 can receive the VPN-IPv4 routes advertised by PE1 and PE2
- PE1 and PE2 can receive the VPN-IPv4 routes advertised by PE3
- Based on the above two conditions, site1 and site3 of VPN1 can intercommunicate, so do the site2 of VPN2 and site3 of VPN1.

PE3 won't advertise VPN-IPv4 routes from PE1 to PE2, or advertise the VPN-IPv4 route from PE2 to PE1 (the routes learnt from an IGBP neighbor won't be sent to other IBGP neighbors), so site1 of VPN1 and site2 of VPN2 can't intercommunicate.

10.3.1.5 BGP/MPLS VPN Route Advertisement

In basic BGP/MPLS VPN networks, VPN route advertisement concerns CE and PE, since P routers only maintains routes of the backbone network, and doesn't need any VPN route information. PE routers only maintain the VPN route information directly connected to it, not all VPN routes. SO the BGP/MPLS VPN network is easy to extend.

The VPN route advertisement process includes three parts to create a reachable route from the local CE to the remote CE, enabling the advertisement of VPN private network route information in the backbone network: from local CE to ingress PE, from the ingress PE to the egress PE, from egress PE to the remote CE.

The followings are introduction to the three parts:

The route information switch from the local CE to the ingress PE

CE will send the local VPN route to the PE directly connected to it after establishing an adjacency to the latter.

CE can use static routes, RIP, OSPF, IS-IS or EBGP to send routes to PE, all in the form of standard IPv4 routes.

The route information switch from the ingress PE to the egress PE

PE will add RD and VPN target attributes to the VPN routes it learns from CE, then store these VPN-IPv4 routes into the VPN instances created for CE.

The ingress PE will advertise the VPN-IPv4 routes to the egress PE via MP-BGP. The egress PE will determine whether to add this route into the route table of VPN instance according to the routes' Export Target attribute and the import Target of the VPN instances it maintains.

Different PE ensure the intercommunication between them via IGP.

The route information switch between the egress PE to the remote CE

Like the route information switch from the local CE to the ingress PE, there are many available methods for the remote CE to learn VPN routes the egress PE, including static route, RIP, OSPF, IS-IS and EBGP.

10.3.1.6 Multi-AS VPN Introduction

In real networking applications, multiple sites of a user VPN may connect to SP with different ASN, or to different AS of the same SP. Such applications of one VPN crossing multiple autonomy systems are called Multi-AS VPN. RFC 2547 provides three Multi-AS VPN resolutions:

- VRF-to-VRF：ASBR use VRF interface to create EBGP neighbors and manage VPN routes, which is also called Inter-Provider Option A;
- EBGP Redistribution of labeled VPN-IPv4 routes： ASBR use MP-EBGP to advertise label VPN-IPv4 routes, which is also called Inter-Provider Option B;
- Multihop EBGP redistribution of labeled VPN-IPv4 routes：PE use Multi-hop MP-EBGP to advertise label VPN-IPv4 routes, which is also called Inter-Provider Option C.

At present we support the first resolution: VRF-to-VRF Multi-VPN resolution.

Multi-VPN resolution

As demonstrated in the nest figure, in this mode, PE routers from two AS directly connects with each other, and serve as ASBR of the AS they belong to. These PE routers (ASBR) connect with each other via VRF interfaces, import all RT this system need the other end to learn, export all RT this system want to obtain from the other, and establish EBGP connections through the VRF interfaces. As a result, the CE they serve will be able to intercommunicate with and isolate from each other like locating in the same AS, with two PE routers treating each other as their own CE. Packets will be forwarded within the AS as VPN packets in the 2-layer label mode, and forwarded as regular IP packets between ASBR.

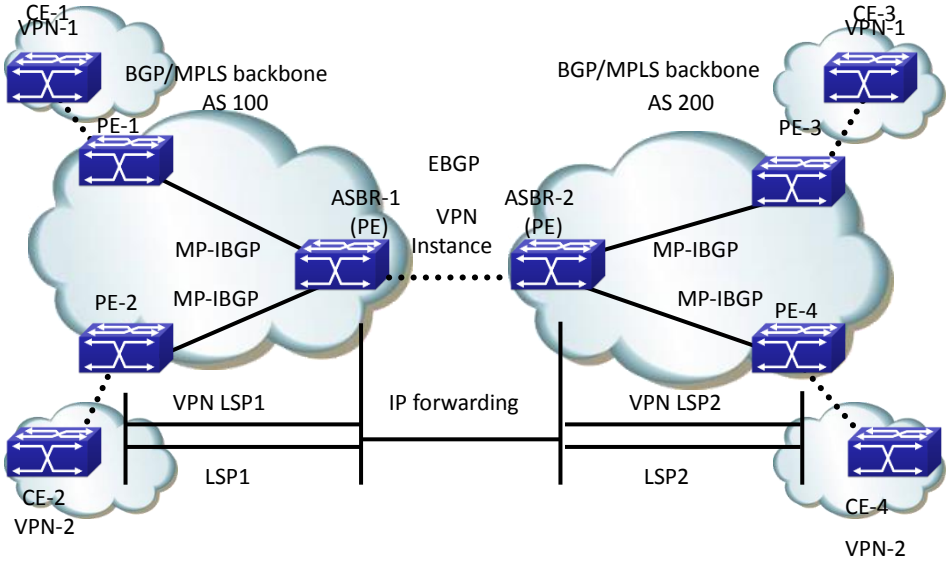


Figure 10-15 Multi-AS VPN Networking

- The advantage of this Multi-AS VPN mode is easy to realize: no special configuration is needed between the two PE serving as ASBR.
- The disadvantage is poor extendibility: the PE serving as ASBR need to manage all VPN routes, and create VPN instances for each VPN. This will cause too many VPN-IPv4 routes on PE.

10.3.2 BGP MPLS VPN Configuration

BGP MPLS VPN configuration task sequence:

Enable globally MPLS (necessary)

Configure VPN instances (necessary)

Create VPN instances, and enter the VPN instance view.

RD Configure the VPN instance RD

Configure the VPN instance RT

Configure the VPN instance to relate with the interface

Configure basic MPLS VPN (necessary)

Configure to use EBGP between PE-CE

Configure the remote PE as the public network VPNv4 neighbor

Enter the BGP-VPN instance view

Configure CE as the VPN private network neighbor

Advertise local private network routes

Configure to use EBGP between PE-CE

Configure the remote PE as the public network VPNv4 neighbor

Create the OSPF instance between PE-CE, and enter the Router OSPF view.

Enable OSPF in the segment between PE-CE

Configure to re-advertise BGP routes

Enter the BGP-VPN instance view

Configure to re-advertise OSPF routes

Advertise local private network routes

RIP Configure to use EBGP between PE-CE

Configure the remote PE as the public network VPNv4 neighbor

Enter the RIP VPN instance view

Enable RIP in the segment between PE-CE

Configure to re-advertise BGP routes

Enter the BGP-VPN instance view

Configure to re-advertise RIP routes

Advertise local private network routes

Configure to use static routes between PE-CE

Configure the remote PE as the public network VPNv4 neighbor

Configure static VPN routes

Enter the BGP-VPN instance view

Configure to re-advertise static routes

Advertise local private network routes

Enable MPLS (necessary)

Command	Explanation
Global Configuration Mode	
mpls enable **no mpls enable**	Necessary Enable MPLS; the no operation will disable MPLS.

Configure VPN instances (necessary)

Create VPN instances and enter VPN instance view

Configure VPN instance RD

Configure VPN instance RT

Configure VPN instance to relate with the interface

Command	Explanation
Global Configuration Mode	
[no] ip vrf <*vrf-name*>	Necessary Create VPN instances; no VPN instance is created by default.
VRF Configuration Mode	

[no] rd <*ASN:nn*_or_*IP-address:nn*>	Necessary Configure VPN instance RD; no RD is created by default.
[no] route-target {import | export | both} <*rt-value*>	Necessary Configure VPN instance RT.
Interface Configuration Mode	
[no] ip vrf forwarding < ***vrf-name*** >	Necessary Configure VPN instance to relate with the interface.
mpls proxy loopback-group <1-max_agg_num> **no mpls proxy loopback-group**	Enable MPLS proxy, the default does not enable the function, enable MPLS proxy when the boardcard of VRF ingress does not support MPLS.
[no] ip address <*ip-address*> <*mask*>	Necessary Configure the private network IP address of the interface directly connecting PE and CE.

3 Configure basic MPLS VPN (necessary)

(1) Configure to use EBGP between PE-CE

Configure the remote PE as the public network VPNv4 neighbor

Enter the BGP-VPN instance view

Configure CE as the VPN private network neighbor

Advertise local private network routes

Command	Explanation
BGP Protocol Configuration Mode	
neighbor <*ip-address*> remote-as <*as-num*>	necessary Configure the remote PE as the public network VPNv4 neighbor. It's suggest to select loopback interface to set up the BGP neighbor among public network PE.
neighbor <*ip-address*> update-source <*as-num*>	Point the local loopback interface for set up neighbor.
Enter the BGP-VPNv4 view	
address-family vpnv4 [unicast]	necessary Create BGP VPNv4. No VPNv4 is created by default.
[no] neighbor <*ip-address*> active	optional Activate all neighbors in VPNv4. All neighbors in VPNv4 view are inactive by default.
BGP Protocol Configuration Mode	
[no] address-family ipv4 {unicast| multicast|vrf <*vrf-nam*>}	optional; Create BGP protocol IPv4 and enter the BGP-VPN instance view. No IPv4 is created by default.

BGP-VPN instance view	
[no] neighbor <*ip-address*> remote-as <*as-num*>	optional Configure CE as the VPN private network neighbor. No private network neighbor is configured by default.
[no] neighbor <*ip-address*> active	optional Activate all neighbors in VPNv4. All neighbors in VPNv4 view are active by default.
[no] redistribute {connected|ospf|rip|static}	optional Configure to re-advertise the directly connected routes and other protocol routes. No re-advertisement of any route by default.

Configure to use EBGP between PE-CE

Configure the remote PE as the public network VPNv4 neighbor

Create the OSPF instance between PE-CE, and enter the Router OSPF view

Enable OSPF in the segment between PE-CE

Configure to re-advertise BGP routes

Enter the BGP-VPN instance view

Configure to re-advertise OSPF routes

Advertise local private network routes

Command	Explanation
BGP Protocol Configuration Mode	
neighbor <*ip-address*> remote-as <*as-num*>	necessary Configure the remote PE as the public network VPNv4 neighbor. It's suggest to select loopback interface to set up the BGP neighbor among public network PE.
neighbor <*ip-address*> update-source <*as-num*>	Point the local loopback interface for set up neighbor.
Enter the BGP-VPNv4 view	
address-family vpnv4 [unicast]	necessary Create BGP VPNv4. No VPNv4 is created by default.
[no] neighbor <*ip-address*> active	optional Activate all neighbors in VPNv4. All neighbors in VPNv4 view are active by default.
Global Configuration Mode	
[no] router ospf [<*process_id*> [<*vrf-nam*>]]	optional Create the OSPF instance between PE-CE, and enter the Router OSPF view.
OSPF VPN instance view	

[no] network {<network> <mask> | <network>/<prefix>} area <area_id>	optioanal Enable OSPF in the segment between PE-CE. Enabled in no segment by default.
[no] redistribute { bgp | connected | static| rip | kernel} [metric-type {1 | 2}] [tag <tag>] [metric <cost_value>] [router-map <WORD>]	optional Configure to re-advertise the BGP routes. No re-advertisement of any route by default.
BGP Protocol Configuration Mode	
[no] address-family ipv4 {unicast| multicast|vrf <*vrf-nam*>}	optional create BGP VPNv4 and enter the BGP-VPN instance view. No VPNv4 is created by default.
BGP-VPN instance view	
[no] redistribute {connected | ospf | rip | static}	optional Configure to re-advertise the directly connected routes and other protocol routes. No re-advertisement of any route by default.

Configure to use EBGP between PE-CE

Configure the remote PE as the public network VPNv4 neighbor

Enter the RIP VPN instance view

Enable RIP in the segment between PE-CE

Configure to re-advertise BGP routes

Enter the BGP-VPN instance view

Configure to re-advertise RIP routes

Advertise local private network routes

Command	Explanation
BGP Protocol Configuration Mode	
neighbor <*ip-address*> remote-as <*as-num*>	necessary Configure the remote PE as the public network VPNv4 neighbor. It's suggest to select loopback interface to set up the BGP neighbor among public network PE.
neighbor <*ip-address*> update-source <*as-num*>	Point the local loopback interface for set up neighbor.
Enter the BGP-VPNv4 view	
address-family vpnv4 [unicast]	necessary Create BGP VPNv4. No VPNv4 is created by default.
[no] neighbor <*ip-address*> active	optional Activate all neighbors in VPNv4. All neighbors in VPNv4 view are active by default.
RIP Protocol Configuration Mode	

[no] address-family ipv4 vrf <*vrf-name*>	optional Create RIP IPv4 protocol family and enter RIP VPN instance view.
RIP VPN instance view	
[no] network {*A.B.C.D/M|ifname|vlan <id>*|loopback <*1-1024*> }	optional Enable the RIP between PE-CE.
[no] redistribute { kernel | connected | static | ospf | isis | bgp} [metric <value>] [route-map<word>]	optional Configure to re-advertise the BGP routes. No re-advertisement of any route by default.
BGP Protocol Configuration Mode	
[no] address-family ipv4 {unicast| multicast|vrf <*vrf-nam*>}	optional Create BGP VPNv4 and enter the BGP-VPN instance view. No VPNv4 is created by default.
BGP-VPN instance view	
[no] redistribute {connected | ospf | rip | static}	optional Configure to re-advertise the directly connected routes and other protocol routes. No re-advertisement of any route by default.

Configure to use static routes between PE-CE

Configure the remote PE as the public network VPNv4 neighbor

Configure static VPN routes

Enter the BGP-VPN instance view

Configure to re-advertise static routes

Advertise local private network routes

Command	Explanation
BGP Protocol Configuration Mode	
neighbor <*ip-address*> remote-as <*as-num*>	necessary Configure the remote PE as the public network VPNv4 neighbor. It's suggest to select loopback interface to set up the BGP neighbor among public network PE.
neighbor <*ip-address*> update-source <*as-num*>	Point the local loopback interface for set up neighbor.
Enter the BGP-VPNv4 view	
address-family vpnv4 [unicast]	necessary Create BGP VPNv4. No VPNv4 is created by default.
[no] neighbor <*ip-address*> active	optional Activate all neighbors in VPNv4. All neighbors in VPNv4 view are active by default.
Global Configuration Mode	

[no] ip route vrf <vrf-name> {<ip-prefix> <mask>|<ip-prefix/<prefix-length>} {<gateway-address>|null0}	optional Manually configure the static VPN routes between PE-CE.
BGP Protocol Configuration Mode	
[no] address-family ipv4 {unicast| multicast|vrf <*vrf-nam*>}	optional Create BGP VPNv4 and enter the BGP-VPN instance view. No VPNv4 is created by default.
BGP-VPN instance view	
[no] redistribute {connected|ospf|rip| static}	optional Configure to re-advertise the static routes, directly connected routes and other protocol routes. No re-advertisement of any route by default.

10.3.3 BGP MPLS VPN Typical Instances

10.3.3.1 Create BGP MPLS VPN between PE-CE via EBGP

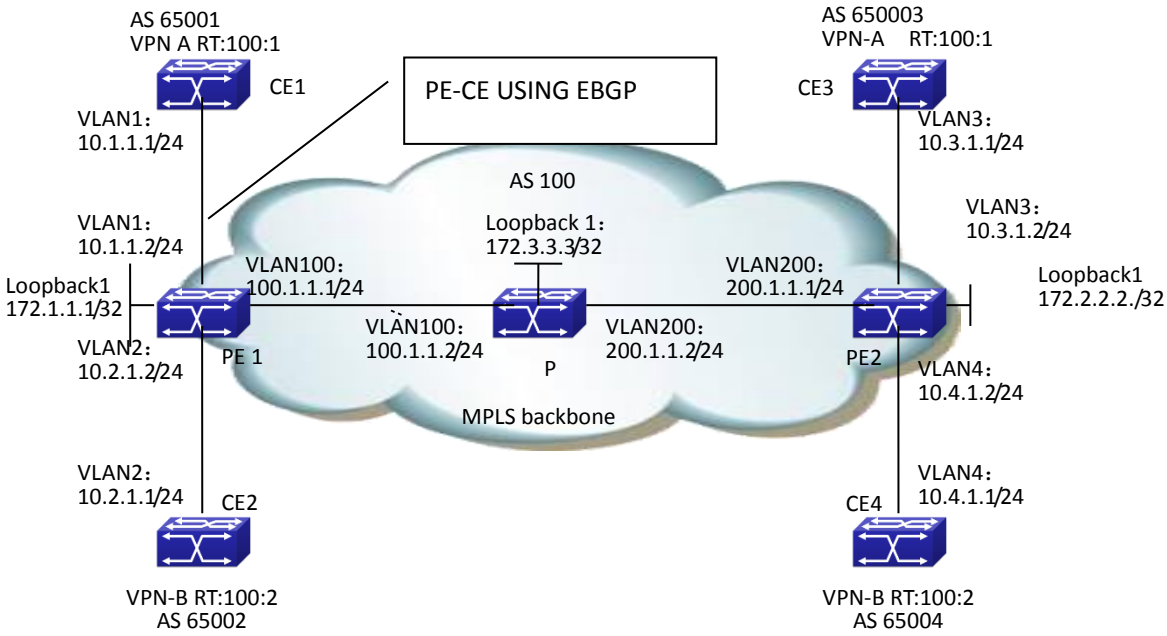


Figure 10-16 Create BGP MPLS VPN between PE-CE via EBGP

The configuration of CE1 is as follows : (the configurations of CE2~CE4 are similar)

CE1#config

```
CE1(config)# interface vlan 1
CE1(config-if-Vlan1)#ip address 10.1.1.1 255.255.255.0
CE1(config-if-Vlan1)#exit
CE1(config)# router bgp 65001
CE1(config-router)#neighbor 10.1.1.2 remote-as 100
CE1(config-router)#redistribute connect
CE1(config-router)#exit
```

The configuration of MPLS BGP on switch PE1 is as follows:

```
Configure VPN instances
PE1#config
PE1(config)#ip vrf vpna
PE1(config-vrf)#rd 100:1
PE1(config-vrf)#route-target both 100:1
PE1(config)#ip vrf vpnb
PE1(config-vrf)#rd 100:2
PE1(config-vrf)#route-target both 100:2

Configure to bind the interface with the VPN instances
PE1(config)# interface vlan 1
PE1(config-if-Vlan1)# ip vrf forwarding vpna
PE1(config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
PE1(config-if-Vlan1)#exit
PE1(config)# interface vlan 2
PE1(config-if-Vlan2)# ip vrf forwarding vpnb
PE1(config-if-Vlan2)#ip address 10.2.1.2 255.255.255.0
PE1(config-if-Vlan2)#exit

Globally enable MPLS and LDP
PE1(config)#mpls enable
PE1(config)#router ldp
PE1(config-router)#exit

LDP Configure the interface and enable LDP
PE1(config)# interface loopback 1
PE1(config-if-Loopback1)# ip address 172.1.1.1 255.255.255.255
PE1(config-if-Loopback1)# exit
PE1(config)# interface vlan 100
PE1(config-if-Vlan100)#ip address 100.1.1.1 255.255.255.0
PE1(config-if-Vlan100)#label-switching
PE1(config-if-Vlan100) #ldp enable
PE1(config-if-Vlan100)#exit

Enable OSPF to advertise the inner network routes
PE1(config)#router ospf
PE1(config-router)# ospf router-id 172.1.1.1
PE1(config-router)# network 0.0.0.0/0 area 0
PE1(config-router)# redistribute connected
```

Configure BGP
PE1(config)# router bgp 100
PE1(config-router)#neighbor 172.2.2.2 remote-as 100
PE1(config-router)#neighbor 172.2.2.2 update-source 172.1.1.1
PE1(config-router)#address-family vpnv4
PE1(config-router-af)#neighbor 172.2.2.2 activate
PE1(config-router-af)#exit
PE1(config-router)# address-family ipv4 vrf vpna
PE1(config-router-af)#neighbor 10.1.1.1 remote-as 65001
PE1(config-router-af)#redistribute connected
PE1(config-router-af)#exit
PE1(config-router)# address-family ipv4 vrf vpnb\
PE1(config-router-af)#neighbor 10.2.1.1 remote-as 65002
PE1(config-router-af)#redistribute connected
PE1(config-router-af)#exit
PE1(config-router)#exit

The configuration of router P is as follows:

Globally enable MPLS and configure LDP on related interfaces.
P#config
P(config)#mpls enable
P(config)#router ldp
P(config-router)#exit
P(config)# interface loopback 1
P(config-if-Loopback1)# ip address 172.3.3.3 255.255.255.255
P(config-if-Loopback1)# exit
P(config)#interface vlan 100
P(config-if-Vlan100)#ip address 100.1.1.2 255.255.255.0
P(config-if-Vlan100)#label-switching
P(config-if-Vlan100)#ldp enable
P(config-if-Vlan100)#exit
P(config)#interface vlan200
P(config-if-Vlan200)#ip address 200.1.1.2 255.255.255.0
P(config-if-Vlan200)#label-switching
P(config-if-Vlan200)#ldp enable
P(config-if-Vlan200)#exit

Configure OSPF
P(config)#router ospf
P(config-router)# ospf router-id 172.3.3.3
P(config-router)# network 0.0.0.0/0 area 0
P(config-router)# redistribute connected

The configuration of switch PE2 is as follows:

Configure VPN instances
PE2#config
PE2(config)#ip vrf vpna

PE2(config-vrf)#rd 100:1
PE2(config-vrf)#route-target both 100:1
PE2(config)#ip vrf vpnb
PE2(config-vrf)#rd 100:2
PE2(config-vrf)#route-target both 100:2

Configure to bind the interface with the VPN instances
PE2(config)# interface vlan 3
PE2(config-if-Vlan3)# ip vrf forwarding vpna
PE2(config-if-Vlan3)#ip address 10.3.1.2 255.255.255.0
PE2(config-if-Vlan3)#exit
PE2(config)# interface vlan 4
PE2(config-if-Vlan4)# ip vrf forwarding vpnb
PE2(config-if-Vlan4)#ip address 10.4.1.2 255.255.255.0
PE2(config-if-Vlan4)#exit

Globally enable MPLS and LDP
PE2(config)#mpls enable
PE1(config)#router ldp
PE1(config-router)#exit

LDP Configure the interface and enable LDP
PE2(config)# interface loopback 1
PE2(config-if-Loopback1)# ip address 172.2.2.2 255.255.255.255
PE2(config-if-Loopback1)# exit
PE2(config)# interface vlan 200
PE2(config-if-Vlan200)#ip address 200.1.1.1 255.255.255.0
PE2(config-if-Vlan200)#label-switching
PE2(config-if-Vlan200) #ldp enable
PE2(config-if-Vlan200)#exit

Enable OSPF to advertise the inner network routes
PE2(config)#router ospf
PE2(config-router)# ospf router-id 172.2.2.2
PE2(config-router)# network 0.0.0.0/0 area 0
PE2(config-router)# redistribute connected

Configure BGP
PE2(config)# router bgp 100
PE2(config-router)#neighbor 172.1.1.1 remote-as 100
PE2(config-router)#neighbor 172.1.1.1 update-source 172.2.2.2
PE2(config-router)#address-family vpnv4
PE2(config-router-af)#neighbor 172.1.1.1 activate
PE2(config-router-af)#exit
PE2(config-router)# address-family ipv4 vrf vpna
PE2(config-router-af)#neighbor 10.3.1.1 remote-as 65003
PE2(config-router-af)#redistribute connected
PE2(config-router-af)#exit

PE2(config-router)# address-family ipv4 vrf vpnb
PE2(config-router-af)#neighbor 10.4.1.1 remote-as 65004
PE2(config-router-af)#redistribute connected
PE2(config-router-af)#exit
PE2(config-router)#exit

10.3.3.2 Create BGP MPLS VPN between PE-CE via OSPF

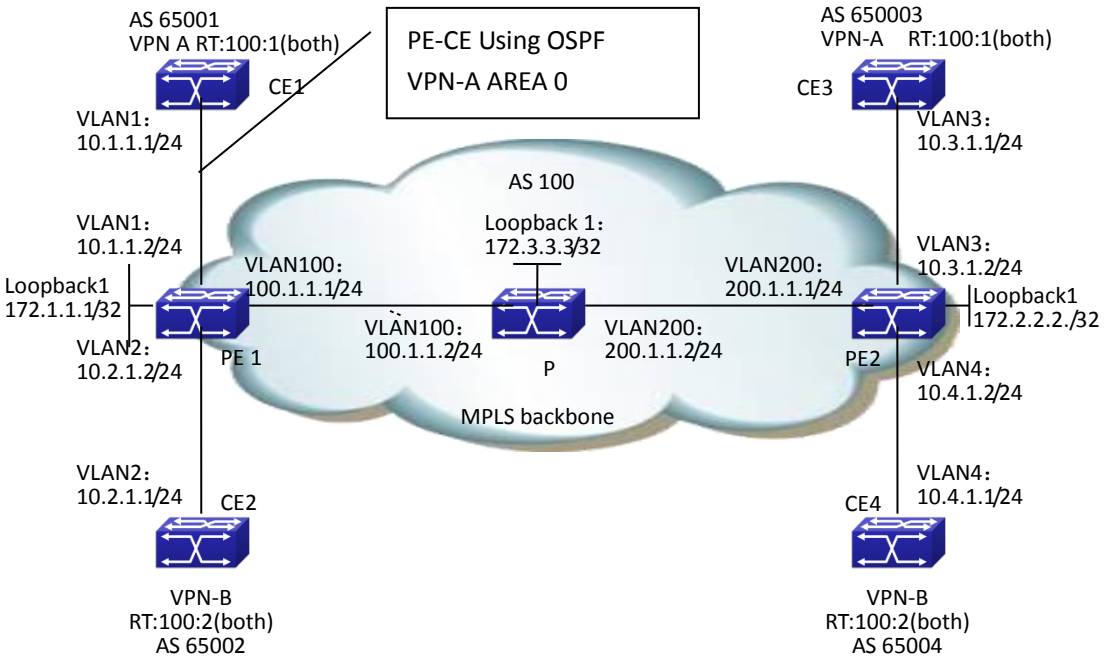


Figure 10-17 Create BGP MPLS VPN between PE-CE via OSPF

The configuration of CE1 is as follows : (the configurations of CE2~CE4 are similar)

CE1#config
CE1(config)# interface vlan 1
CE1(config-if-Vlan1)#ip address 10.1.1.1 255.255.255.0
CE1(config-if-Vlan1)#exit
CE1(config)# router ospf
CE1(config-router)#network 0.0.0.0/0 area 0
CE1(config-router)#redistribute connect
CE1(config-router)#exit

The confiugraiton of MPLS BGP on switch PE1 is as follows : (the configuration of PE2 is similar)

Configure VPN instances
PE1#config
PE1(config)#ip vrf vpna
PE1(config-vrf)#rd 100:1

PE1(config-vrf)#route-target both 100:1
PE1(config)#ip vrf vpnb
PE1(config-vrf)#rd 100:2
PE1(config-vrf)#route-target both 100:2

Configure to bind the interface with the VPN instances
PE1(config)# interface vlan 1
PE1(config-if-Vlan1)# ip vrf forwarding vpna
PE1(config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
PE1(config-if-Vlan1)#exit
PE1(config)# interface vlan 2
PE1(config-if-Vlan2)# ip vrf forwarding vpnb
PE1(config-if-Vlan1)#ip address 10.2.1.2 255.255.255.0
PE1(config-if-Vlan1)#exit

Globally enable MPLS and LDP
PE1(config)#mpls enable
PE1(config)#router ldp
PE1(config-router)#exit

LDP Configure the interface and enable LDP
PE1(config)# interface loopback 1
PE1(config-if-Loopback1)# ip address 172.1.1.1 255.255.255.255
PE1(config-if-Loopback1)# exit
PE1(config)# interface vlan 100
PE1(config-if-Vlan100)#ip address 100.1.1.1 255.255.255.0
PE1(config-if-Vlan100)# label-switching
PE1(config-if-Vlan100) #ldp enable
PE1(config-if-Vlan100)#exit

Enable OSPF to advertise the inner network routes
PE1(config)#router ospf
PE1(config-router)# ospf router-id 172.1.1.1
PE1(config-router)# network 0.0.0.0/0 area 0
PE1(config-router)# redistribute connected
PE1(config-router)#exit

Enable OSPF VRF to advertise the private network routes
PE1(config)#router ospf 1 vpna
PE1(config-router)# network 0.0.0.0/0 area 0
PE1(config-router)#redistribute connected
PE1(config-router)#redistribute bgp
PE1(config-router)#exit
PE1(config)#router ospf 1 vpnb
PE1(config-router)# network 0.0.0.0/0 area 0
PE1(config-router)#redistribute connected
PE1(config-router)#redistribute bgp
PE1(config-router)#exit

Configure BGP
PE1(config)# router bgp 100
PE1(config-router)#neighbor 172.2.2.2 remote-as 100
PE1(config-router)#neighbor 172.2.2.2 update-source 172.1.1.1
PE1(config-router)#address-family vpnv4
PE1(config-router-af)#neighbor 172.2.2.2 activate
PE1(config-router-af)#exit
PE1(config-router)# address-family ipv4 vrf vpna
PE1(config-router-af)#redistribute connected
PE1(config-router-af)#redistribute ospf
PE1(config-router-af)#exit
PE1(config-router)# address-family ipv4 vrf vpnb
PE1(config-router-af)#redistribute connected
PE1(config-router-af)#redistribute ospf
PE1(config-router-af)#exit
PE1(config-router)#exit

The configuration of router P is as follows:

Globally enable MPLS and configure LDP on related interfaces.
P#config
P(config)#mpls enable
P(config)#router ldp
P(config-router)#exit
P(config)# interface loopback 1
P(config-if-Loopback1)# ip address 172.3.3.3 255.255.255.255
P(config-if-Loopback1)# exit
P(config)#interface vlan 100
P(config-if-Vlan100)#ip address 100.1.1.2 255.255.255.0
P(config-if-Vlan100)#label-switching
P(config-if-Vlan100)#ldp enable
P(config-if-Vlan100)#exit
P(config)#interface vlan200
P(config-if-Vlan200)#ip address 200.1.1.2 255.255.255.0
P(config-if-Vlan200)#label-switching
P(config-if-Vlan200)#ldp enable
P(config-if-Vlan200)#exit

Configure OSPF
P(config)#router ospf
P(config-router)# ospf router-id 172.3.3.3
P(config-router)# network 0.0.0.0/0 area 0
P(config-router)# redistribute connected

10.3.3.3 Create BGP MPLS VPN between PE-CE via RIP

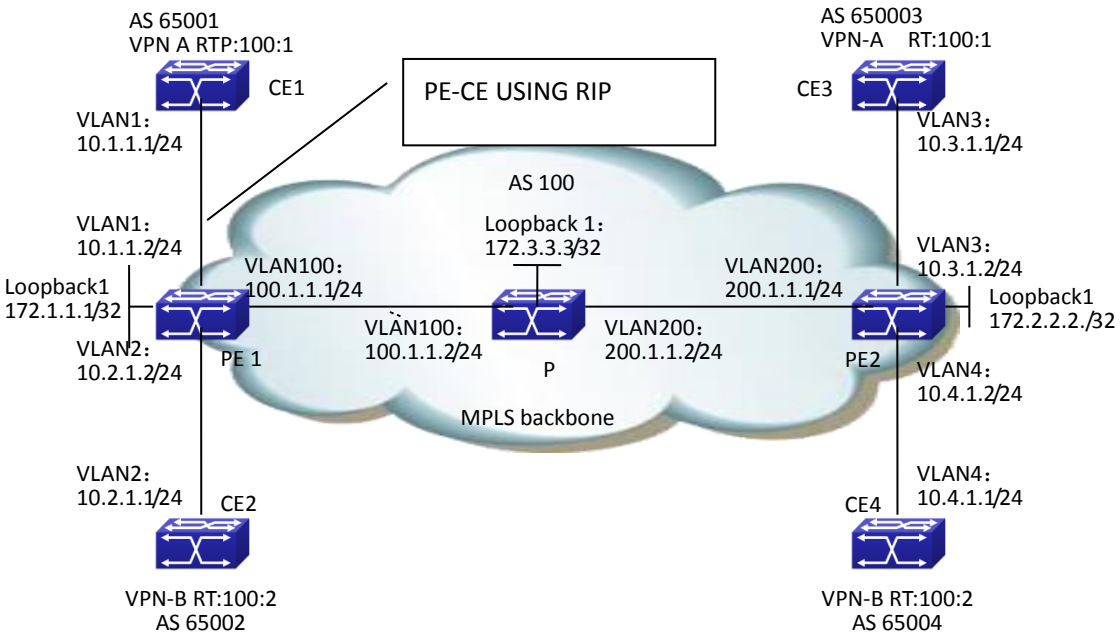


Figure 10-18 Create BGP MPLS VPN between PE-CE via RIP

The configuration of CE1 is as follows : (the configurations of CE2~CE4 are similar)

CE1#config
CE1(config)# interface vlan 1
CE1(config-if-Vlan1)#ip address 10.1.1.1 255.255.255.0
CE1(config-if-Vlan1)#exit
CE1(config)# router rip
CE1(config-router)#network 0.0.0.0/0
CE1(config-router)#redistribute connect
CE1(config-router)#exit

The confiugraiton of MPLS BGP on switch PE1 is as follows : (the configuration of PE2 is similar)

Configure VPN instances
PE1#config
PE1(config)#ip vrf vpna
PE1(config-vrf)#rd 100:1
PE1(config-vrf)#route-target both 100:1
PE1(config)#ip vrf vpnb
PE1(config-vrf)#rd 100:2
PE1(config-vrf)#route-target both 100:2

Configure to bind the interface with the VPN instances
PE1(config)# interface vlan 1
PE1(config-if-Vlan1)# ip vrf forwarding vpna
PE1(config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
PE1(config-if-Vlan1)#exit
PE1(config)# interface vlan 2

PE1(config-if-Vlan2)# ip vrf forwarding vpnb
PE1(config-if-Vlan1)#ip address 10.2.1.2 255.255.255.0
PE1(config-if-Vlan1)#exit

Globally enable MPLS and LDP
PE1(config)#mpls enable
PE1(config)#router ldp
PE1(config-router)#exit

LDP Configure the interface and enable LDP
PE1(config)# interface loopback 1
PE1(config-if-Loopback1)# ip address 172.1.1.1 255.255.255.255
PE1(config-if-Loopback1)# exit
PE1(config)# interface vlan 100
PE1(config-if-Vlan100)#ip address 100.1.1.1 255.255.255.0
PE1(config-if-Vlan100)#label-switching
PE1(config-if-Vlan100) #ldp enable
PE1(config-if-Vlan100)#exit

Enable OSPF to advertise the inner network routes
PE1(config)#router ospf
PE1(config-router)# ospf router-id 172.1.1.1
PE1(config-router)# network 0.0.0.0/0 area 0
PE1(config-router)# redistribute connected
PE1(config-router)#exit

Enable OSPF VRF to advertise the private network routes
PE1(config)#router rip
PE1(config-router)#address-family ipv4 vrf vpna
PE1(config-router-af)#network 0.0.0.0/0
PE1(config-router-af)#redistribute connected
PE1(config-router-af)#redistribute bgp
PE1(config-router-af)#exit
PE1(config-router)#address-family ipv4 vrf vpnb
PE1(config-router-af)#network 0.0.0.0/0
PE1(config-router-af)#redistribute connected
PE1(config-router-af)#redistribute bgp
PE1(config-router-af)#exit
PE1(config-router)#exit

Configure BGP
PE1(config)# router bgp 100
PE1(config-router)#neighbor 172.2.2.2 remote-as 100
PE1(config-router)#neighbor 172.2.2.2 update-source 172.1.1.1
PE1(config-router)#address-family vpnv4
PE1(config-router-af)#neighbor 172.2.2.2 activate
PE1(config-router-af)#exit
PE1(config-router)# address-family ipv4 vrf vpna

PE1(config-router-af)#redistribute connected
PE1(config-router-af)#redistribute ospf
PE1(config-router-af)#exit
PE1(config-router)# address-family ipv4 vrf vpnb
PE1(config-router-af)#redistribute connected
PE1(config-router-af)#redistribute ospf
PE1(config-router-af)#exit
PE1(config-router)#exit

The configuration of switch P is as follows

Globally enable MPLS and configure LDP on related interfaces.
P#config
P(config)#mpls enable
P(config)#router ldp
P(config-router)#exit
P(config)# interface loopback 1
P(config-if-Loopback1)# ip address 172.3.3.3 255.255.255.255
P(config-if-Loopback1)# exit
P(config)#interface vlan 100
P(config-if-Vlan100)#ip address 100.1.1.2 255.255.255.0
P(config-if-Vlan100)#label-switching
P(config-if-Vlan100)#ldp enable
P(config-if-Vlan100)#exit
P(config)#interface vlan200
P(config-if-Vlan200)#ip address 200.1.1.2 255.255.255.0
P(config-if-Vlan200)#label-switching
P(config-if-Vlan200)#ldp enable
P(config-if-Vlan200)#exit

Configure OSPF
P(config)#router ospf
P(config-router)# ospf router-id 172.3.3.3
P(config-router)# network 0.0.0.0/0 area 0
P(config-router)# redistribute connected

10.3.3.4 Create BGP MPLS VPN between PE-CE via Static Routes

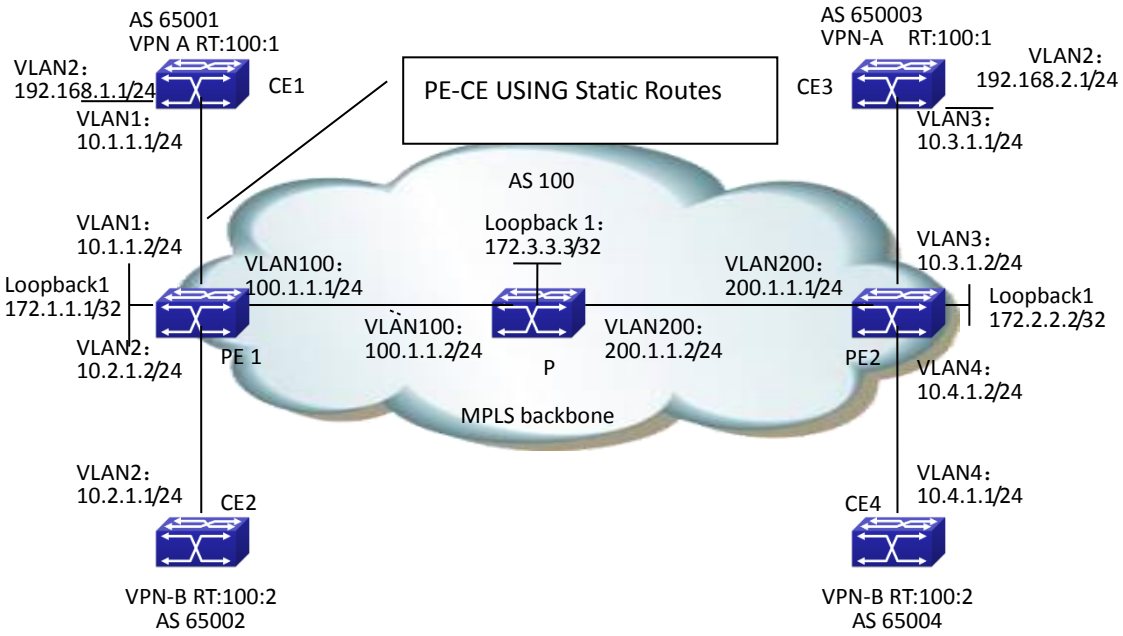


Figure 10-19 Create BGP MPLS VPN between PE-CE via Static Routes

The configuration of CE1 is as follows : (the configurations of CE2~CE4 are similar)

CE1#config
CE1(config)# interface vlan 1
CE1(config-if-Vlan1)#ip address 10.1.1.1 255.255.255.0
CE1(config-if-Vlan1)#exit
CE1(config)# interface loopback 1
CE1(config-if-Vlan1)#ip address 192.168.1.1 255.255.255.0
CE1(config-if-Vlan1)# exit
CE1(config)# ip route vrf vpna 192.168.2.1/24 10.1.1.2

The confiugraiton of MPLS BGP on switch PE1 is as follows : (the configuration of PE2 is similar)

Configure VPN instances
PE1#config
PE1(config)#ip vrf vpna
PE1(config-vrf)#rd 100:1
PE1(config-vrf)#route-target both 100:1
PE1(config)#ip vrf vpnb
PE1(config-vrf)#rd 100:2
PE1(config-vrf)#route-target both 100:2

Configure to bind the interface with the VPN instances
PE1(config)# interface vlan 1
PE1(config-if-Vlan1)# ip vrf forwarding vpna
PE1(config-if-Vlan1)#ip address 10.1.1.2 255.255.255.0
PE1(config-if-Vlan1)#exit
PE1(config)# interface vlan 2
PE1(config-if-Vlan2)# ip vrf forwarding vpnb

PE1(config-if-Vlan1)#ip address 10.2.1.2 255.255.255.0
PE1(config-if-Vlan1)#exit

Globally enable MPLS and LDP
PE1(config)#mpls enable
PE1(config)#router ldp
PE1(config-router)#exit

Configure the interface and enable LDP
PE1(config)# interface loopback 1
PE1(config-if-Loopback1)# ip address 172.1.1.1 255.255.255.255
PE1(config-if-Loopback1)# exit
PE1(config)# interface vlan 100
PE1(config-if-Vlan100)#ip address 100.1.1.1 255.255.255.0
PE1(config-if-Vlan100) #ldp enable
PE1(config-if-Vlan100)#exit

Enable OSPF to advertise the inner network routes
PE1(config)#router ospf
PE1(config-router)# ospf router-id 172.1.1.1
PE1(config-router)# network 0.0.0.0/0 area 0
PE1(config-router)# redistribute connected
PE1(config-router)#exit

Configure static private network routes
PE1(config)# ip route vrf vpna 192.168.1.1/24 10.1.1.2
PE1(config)# ip route vrf vpnb 192.168.2.1/24 10.1.1.2
PE1(config-router)#address-family ipv4 vrf vpna
PE1(config-router-af)#network 0.0.0.0/0
PE1(config-router-af)#redistribute connected
PE1(config-router)#exit

Configure BGP
PE1(config)# router bgp 100
PE1(config-router)#neighbor 172.2.2.2 remote-as 100
PE1(config-router)#neighbor 172.2.2.2 update-source 172.1.1.1
PE1(config-router)#address-family vpnv4
PE1(config-router-af)#neighbor 172.2.2.2 activate
PE1(config-router-af)#exit
PE1(config-router)# address-family ipv4 vrf vpna
PE1(config-router-af)#redistribute connected
PE1(config-router-af)#redistribute static
PE1(config-router-af)#exit
PE1(config-router)# address-family ipv4 vrf vpnb
PE1(config-router-af)#redistribute connected
PE1(config-router-af)# redistribute static
PE1(config-router-af)#exit
PE1(config-router)#exit

The configuration of switch P is as follows

Globally enable MPLS and configure LDP on related interfaces.

```
P#config
P(config)#mpls enable
P(config)#router ldp
P(config-router)#exit
P(config)# interface loopback 1
P(config-if-Loopback1)# ip address 172.3.3.3 255.255.255.255
P(config-if-Loopback1)# exit
P(config)#interface vlan 100
P(config-if-Vlan100)#ip address 100.1.1.2 255.255.255.0
P(config-if-Vlan100)#ldp enable
P(config-if-Vlan100)#exit
P(config)#interface vlan200
P(config-if-Vlan200)#ip address 200.1.1.2 255.255.255.0
P(config-if-Vlan100)#ldp enable
P(config-if-Vlan200)#exit
```

Configure OSPF

```
P(config)#router ospf
P(config-router)# ospf router-id 172.3.3.3
P(config-router)# network 0.0.0.0/0 area 0
P(config-router)# redistribute connected
```

10.3.4 MPLS BGP VPN Troubleshooting

When configuring and using MPLS BGP VPN, some problems like incorrect physical connections, configuration errors may cause it to fail, so please pay attention to the following notices to avoid them:

- First, make sure the creation of OSPF neighbors between PE1, P and PE2, the advertisement of routes including the loopback interface and the creation of BGP neighbor between PE are correct.
- Second, make sure the LDP is globally enabled on PE1, P and PE2, and correctly enabled on active interfaces. Check whether the establishment of LDP sessions on PE1, P and PE2 is correct.
- Then, make sure the PE-CE route advertisement mode used when creating the VPN and corresponding configuration are correct. Check whether CE advertises related private network route to the remote PE. Please notice that CE needs no VRF instance. If EBGP is used to advertise the private network routes, the BGP ASN between CE1 and CE2 shouldn't be the same, or the loop detection of BGP will filter the corresponding private network routes.
- Next, make sure the BGP VPN instances on PE are correctly configured. When using OSPF or RIP to create and advertise PE-CE routes, please import BGP

routs and import corresponding OSPF and RIP routes to the BGP VPN instances. Implementing “show ip bgp vpnv4 all” on PE1 will display the route information of CE1 and CE2, if the configuration is correct. Implementing “show mpls vrf-table” on PE will display that the labels are distributed to corresponding private network routes, and the state is UP. If the Oper status in the vrf-table of the corresponding private network routes, use “show mpls ftn-table” to check whether the corresponding FEC create ftn.

- ☞ At last, if all above steps are correct, use “show ip route” on CE1 and CE2 to check the correct route information in the VPN. It is not recommended for users to create VPN via the static routes unless very familiar with BGP MPLS VPN.
- ☞ Besides, if no remote CE device can be checked on CE after saving the correction configuration and rebooting the device, please be patience, since the establishing OSPF, LDP, BGP connections and advertising routes are time-consuming.

10.4 Public Network Access of MPLS VPN

10.4.1 Public Network Access Introduction

Public network access of VPN means the ability of VPN sites to access public Internet. RFC4364 defines the basic protocol regulations, including some methods for VPN to access Internet:

- ☞ Non-VRF Internet Access Mode
- ☞ VRF Internet Access Mode 1
- ☞ VRF Internet Access Mode 3

10.4.1.1 Non-VRF Internet Access Mode

As demonstrated in the next figure, in non-VRF Internet Access Mode, PE routers communicate with Internet gateways via non-VFP interface; and the Internet access traffic of VPN sites are forwarded according to the global route table of PE routers. The CE and PE routers capable of accessing Internet have two connections, one with the public network interface of PE (public network connection), the other with the priviate network interface of PE (private network connection). The global route table of PE routes can contain the whole or part of Internet routes, or only a default routes pointing to the Internet gateway. CE routers learn Interet routs via the public network connection, and advertise to PE via the public network connection the globally registered IP address sub-net routes in the VPN site, which will be advertised to the Internet gateway by PE and finally to Internet.

The Internet access traffic of VPN sites is also sent and received by the public network connection. The private network connection between CE and PE is for the route learning of CE and advertising the private network routes in tehe VPN. The VPN sites also communicate via private network connections, and forward according to the VRF route table of PE routers. In this mode, the global and VRF rout table of PE routers are completely isolated ; and the distribution of VPN routes and Internet routes are completely independent.

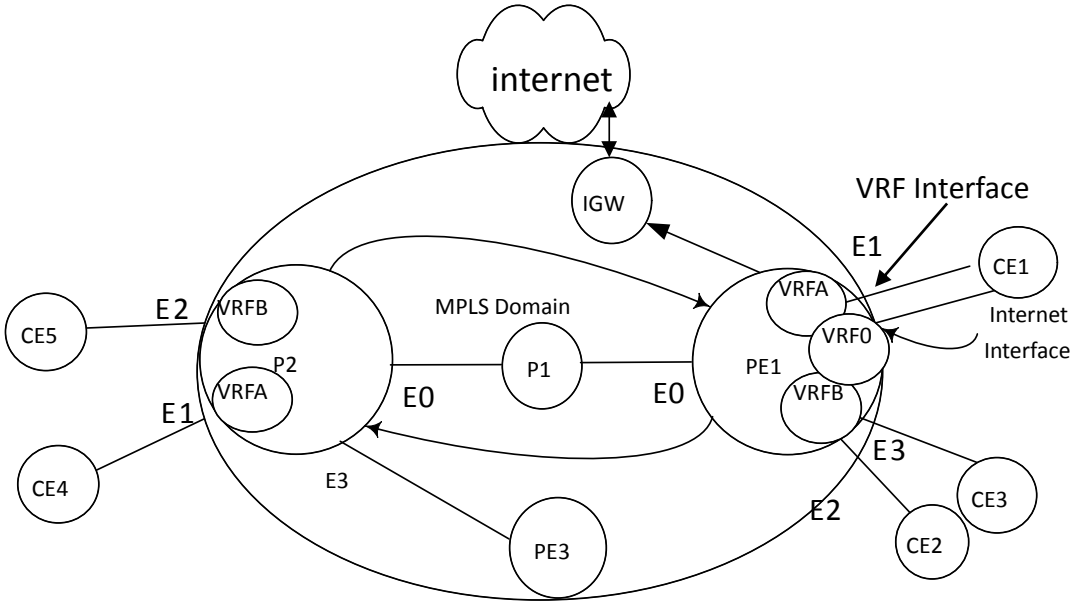


Fig 10-20 Non-VRF Internet Access Mode

10.4.1.2 VRF Internet Access Mode 1

As demonstrated in the next figure, in VRF Internet Access Mode 1, PE routers communicate with Internet gateways via non-VFP interface. The Internet access traffic of VPN sites and the traffic between VPN sites are sent and received via the private netwrok connections between CE and PE. PE routes contain the whole or part of Internet routes, or only a default routes pointing to the Internet gateway. When the IP packets accessing Internet from VPN reach the VRF interfaces of PE, a failed lookup in the VRF route table will cause a lookup in the global route table. If a match is found, the pakce will be forwarded to the Internet gateway, and finaly to Internet via the gateway. To emable the Internet hosts access VPN sites, a special static route needs to be registered in the PE global route table, whose destiantion segment is the IP address sub-net address which is globally registered in the VPN site, egress interface is the private network interface pointing to the VPN site, and next-hop is CE router. This static route is advertised to the internet gateway by PE, and then to Internet by the gateway. When the packets from the Internet to the VPN reach the pbulic network interface of PE, it will be forwarded to the next-hop via the private network interface if it matches the static route in the PE's global route tabel poinging to the VPN site. In this mode, the global route table and VRF route tabel of PE routers are not completely isolated, since the global one contains part of VPN routes.

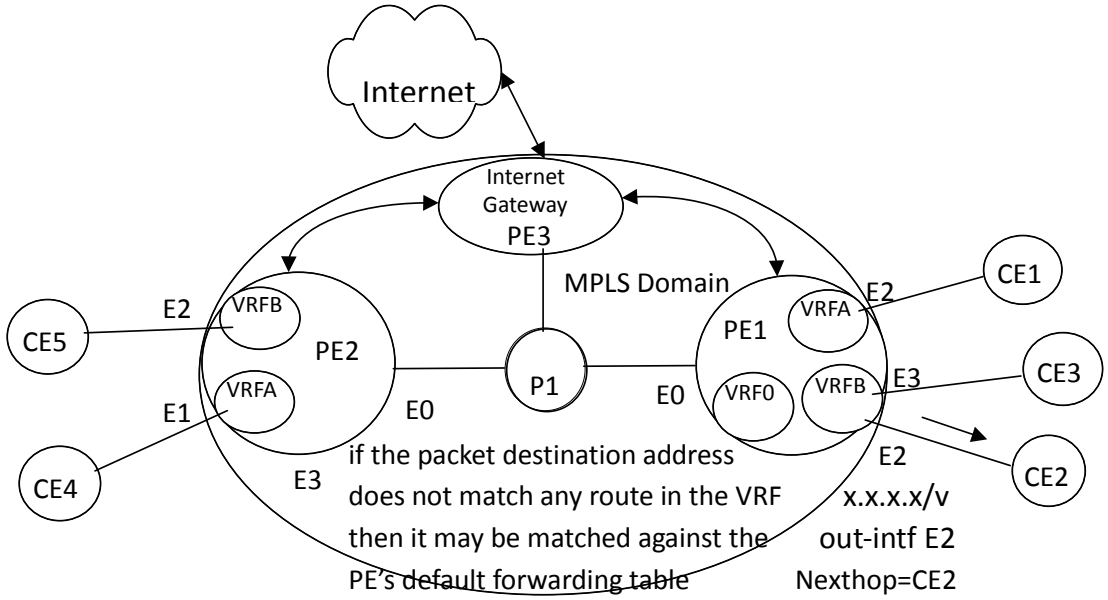


Fig 10-21 VRF Internet Access Mode 1

10.4.1.3 VRF Internet Access Mode 3

In VRF Internet Access Mode 3, as demonstrated in the next figure, VPN site access the Internet via private network connections between PE and CE. The VRF route tabl eof PE routers contain Internet routes, which are learnt via the PE routers conencted with the Internet gateway (Internet PE). Internet PE will create an Internet VRF, and connect withe Internet gateway withe the interface bound with the Internet VRF. Thus , the Internet gateway will be able to advertise Inernet routes to the VRF route table of the Internet PE. There routes then will be advertised to VRF of other PE routers as VPNv4 routes. PE routers connected with the VPN sites needing to access the internet will advertise correspondng VRF routes (only those routes whose destiatnion segments are globally registered IP addess sub-net in the VPN) to Internet PE via VPNv4 route. These routes will be added into the Internet VRF and then advertised to Internet by the Internet gateway. The import and export policy of these routes depend on the route-target configuration of MBGP and vrf. Please notice that, in this mode, no overlap of address or route is allowed between the VPN sites capable of accessing the Internet.

It is not recommended for users to access the public network in this mode, for a large number of Internet routes will be imported to PE.

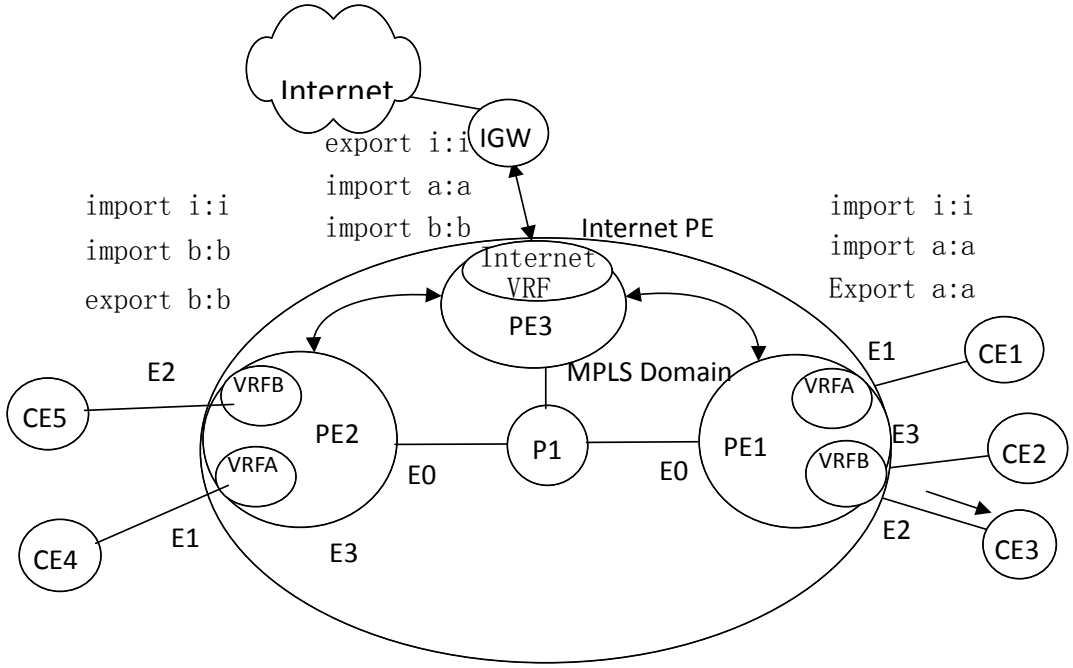


Fig 10-22 VRF Internet Access Mode 3

10.4.2 Public Network Access Configuration

Public Network Access Configuration Task Sequence:

Configure non-VRP Internet access mode

Configure regular L3VPN

Add a public connection between CE and PE, the connection interface is a non-VRF one.

Filter routes on CE; advertise public network routes to PE via the public network connection.

Configure proper filter policy on the public network interface, to filter the packets whose source and destination addresses are private network addresses.

Configure default routes

IGW import the default routes to BGP

PE advertise the default routes to CE via the public network connection

CE advertise the default routes to PE via the private network connection, and then to other CE.

Configure the static route

Configure the static route pointing to Internet on CE1

Configure the static route pointing to the public network interface of CE on PE1

Configure VRP Internet access mode 1

Configure regular L3VPN

Configure ip vrf forwarding VPNA fallback global on the private network interface of PE

Configure 3 static routes:

Configure a default route on CE, whose next-hop is the proxy server

Add a default route to Internet on PE, whose next-hop is IGW. PE advertises a default route via OSPF, whose next-hop is the PE itself.

Add a static route form Internet to proxy server to the global route table of PE, whose destination is VPN public network address, next-hop is proxy server; and advertise this route to other PE via OSPF

Configure non-VRF Internet Access Mode

This configuration concerns no extra command line other than the configuration sequence. Please refer to the configuration instruction of the corresponding function for details about commands

Configure VRP Internet access mode 1

Configure VRP Internet access mode 1

Configure regular L3VPN

Configure ip vrf forwarding VPNA fallback global on the private network interface of PE

Configure 3 static routes

Configure a default route on CE, whose next-hop is the proxy server

Add a default route to Internet on PE, whose next-hop is IGW. PE advertises a default route via OSPF, whose next-hop is the PE itself.

Add a static route form Internet to proxy server to the global route table of PE, whose destination is VPN public network address, next-hop is proxy server; and advertise this route to other PE via OSPF

Command	Explanation
Configure regular L3VPN	Refer to the BGP MPLS VPN configuration
Interface Configuration Mode	
[no] ip vrf forwarding <*vrf_name*> fallback global	Necessary Configure the global second lookup function of VRF route table. It is not configured by default. Before this configuration, cancel the VRF configuration in the interface view.
Global Configuration Mode	
[no] ip route vrf <*vrf-name*> {<ip-prefix> <mask>|<ip-prefix/<prefix-length>} {<gateway-address>|null0}	Necessary Configure static routes, Only three are needed: one is the default route of CE1, another is the default route to Internet on PE3 and the other is the static route form Internet to the proxy server on PE1.

10.4.3 Public Network Access Typical Instances

10.4.3.1 Non-VRF Internet Access Mode

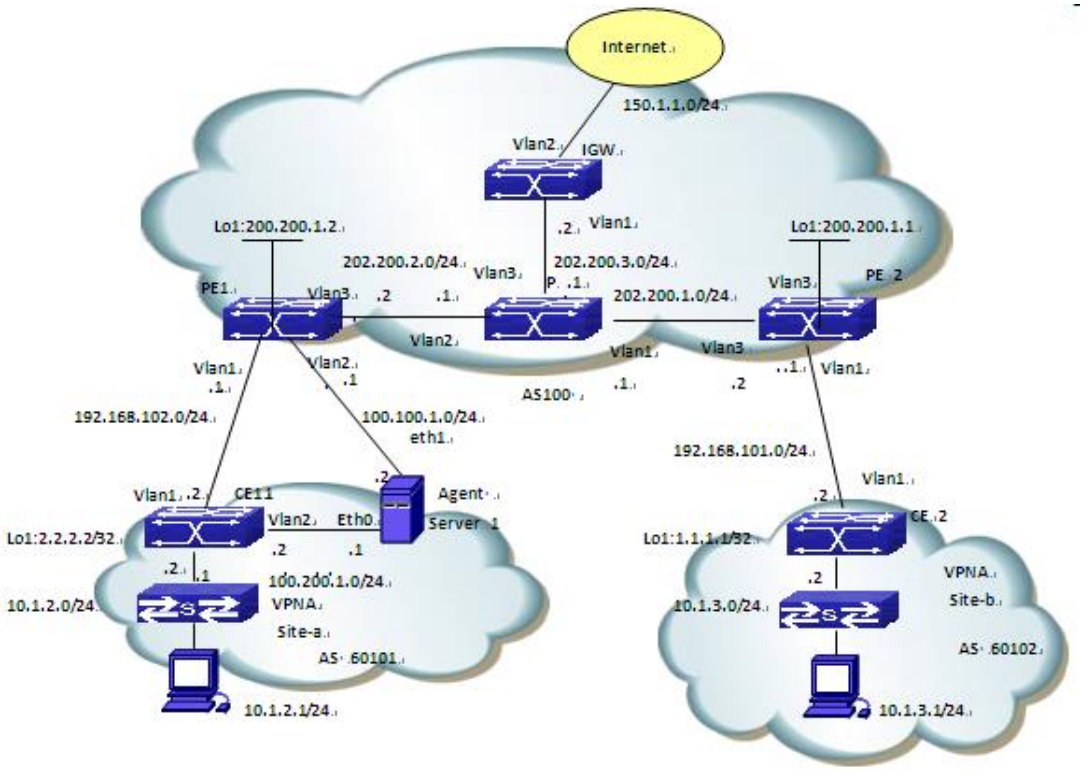


Fig 10-23 Non-VRF Internet Access Mode

The configuration of CE1 is as follows:

```
CE1#config
CE1(config)#access-list 1 deny 100.100.1.0 0.0.0.255
CE1(config)#access-list 1 deny 100.200.1.0 0.0.0.255
CE1(config)#access-list 1 permit any-source
CE1(config)#access-list 2 permit 10.1.1.0 0.0.0.255
CE1(config)#access-list 2 permit 10.1.2.0 0.0.0.255
CE1(config)#access-list 2 deny any-source
CE1(config)# interface vlan 1
CE1(config-if-Vlan1)#ip address 192.168.102.2 255.255.255.0
CE1(config-if-Vlan1)#exit
CE1(config)# interface vlan 2
CE1(config-if-Vlan1)#ip address 100.200.1.2 255.255.255.0
CE1(config-if-Vlan1)#exit
CE1(config)# interface vlan 3
CE1(config-if-Vlan1)#ip address 10.1.2.2 255.255.255.0
CE1(config-if-Vlan1)#exit
CE1(config)# interface loopback 1
CE1(config-if-Vlan1)#ip address 2.2.2.2 255.255.255.255
CE1(config-if-Vlan1)# exit
CE1(config)#router bgp 60102
```

```
CE1(config-router)#network 120.1.1.0/24
CE1(config-router)#network 120.1.2.0/24
CE1(config-router)#network 10.1.2.0/24
CE1(config-router)#redistribute connected
CE1(config-router)#neighbor 100.100.1.1 remote-as 100
CE1(config-router)#neighbor 100.100.1.1 distribute-list 2 out
CE1(config-router)#neighbor 192.168.102.1 remote-as 100
CE1(config-router)#neighbor 192.168.102.1 default-originate
CE1(config-router)#neighbor 192.168.102.1 distribute-list 1 out
CE1(config-router)#exit
CE1(config)# ip route 100.100.1.1 255.255.255.0 100.200.1.1
CE1(config)# ip route 0.0.0.0/0 100.200.1.1
CE1(config)# exit
```

The configuration of PE1 is as follows:

```
PE1#config
PE1(config)#access-list 100 deny ip 10.1.2.0 0.0.0.255 any-destination
PE1(config)#access-list 100 deny ip 10.1.2.0 0.0.0.255 any-destination
PE1(config)#access-list 100 deny ip 10.1.3.0 0.0.0.255 any-destination
PE1(config)#access-list 100 deny ip anysource 200.200.1.0 0.0.0.255
PE1(config)#access-list 100 deny ip anysource 202.200.0.0 0.0.255.255
PE1(config)#firewall enable
PE1(config-vrf)#ip vrf VRF-A
PE1(config-vrf)#rd 100:10
PE1(config-vrf)#route-target both 100:10
PE1(config-vrf)#exit
PE1(config)#interface vlan1
PE1(config-if-Vlan1)#ip vrf forwarding VRF-A
PE1(config-if-Vlan1)#ip address 192.168.102.1 255.255.255.0
PE1(config-if-Vlan1)#exit
PE1(config)#interface vlan2
PE1(config-if-Vlan2)#ip address 100.100.1.1 255.255.255.0
PE1(config-if-Vlan2)#ip access-group 1 in
PE1(config-if-Vlan2)#exit
PE1(config)# interface vlan3
PE1(config-if-Vlan3)#label-switching
PE1(config-if-Vlan3)#ldp enable
PE1(config-if-Vlan3)#ip address 202.200.2.2 255.255.255.0
PE1(config-if-Vlan3)#exit
PE1(config)#interface Loopback1
PE1(config)#ip address 200.200.1.2 255.255.255.255
PE1(config)#router ospf
PE1(config-router)#network 200.200.1.2/32 area 0
PE1(config-router)#network 202.200.2.0/24 area 0
PE1(config-router)#exit
```

```
PE1(config)#router bgp 100
PE1(config-router)#neighbor 100.200.1.2 remote-as 60102
PE1(config-router)#neighbor 200.200.1.1 remote-as 100
PE1(config-router)#neighbor 202.200.3.2 remote-as 100
PE1(config-router)#neighbor 202.200.3.2 next-hop-self
PE1(config-router)#address-family vpnv4 unicast
PE1(config-router-af)#neighbor 200.200.1.1 activate
PE1(config-router-af)#exit-address-family
PE1(config-router)#address-family ipv4 vrf VRF-A
PE1(config-router-af)#neighbor 192.168.102.2 remote-as 60102
PE1(config-router-af)#no neighbor 192.168.102.2 send-community extended
PE1(config-router-af)#exit-address-family
PE1(config-router)#exit
PE1(config)# router ldp
PE1(config-router)#ip route 100.200.1.2 255.255.255.0 100.100.1.2
```

The configuration of P is as follows:

```
P#config
P(config)#interface Vlan1
P(config-if-Vlan1)#label-switching
P(config-if-Vlan1)#ldp enable
P(config-if-Vlan1)#ip address 202.200.1.1 255.255.255.0
P(config-if-Vlan1)#exit
P(config)#interface Vlan2
P(config-if-Vlan2)#label-switching
P(config-if-Vlan2)#ldp enable
P(config-if-Vlan2)#ip address 202.200.2.1 255.255.255.0
P(config-if-Vlan2)#exit
P(config)#interface Vlan3
P(config-if-Vlan3)#ip address 202.200.3.1 255.255.255.0
P(config-if-Vlan3)#exit
P(config)#router ospf
P(config-router)#network 202.200.1.0/24 area 0
P(config-router)#network 202.200.2.0/24 area 0
P(config-router)#network 202.200.3.0/24 area 0
P(config-router)#exit
P(config)#router ldp
```

The configuration of PE2 is as follows:

```
PE2#config
PE2(config)#ip vrf VRF-A
PE2(config-vrf)#rd 100:10
PE2(config-vrf)#route-target both 100:10
PE2(config-vrf)#exit
PE2(config)#interface Vlan1
PE2(config-if-Vlan1)#ip vrf forwarding VRF-A
```

```
PE2(config-if-Vlan1)#ip address 192.168.101.1 255.255.255.0
PE2(config-if-Vlan1)#exit
PE2(config)#interface Vlan2
PE2(config-if-Vlan2)#label-switching
PE2(config-if-Vlan2)#ldp enable
PE2(config-if-Vlan2)#ip address 202.200.1.2 255.255.255.0
PE2(config-if-Vlan2)#exit
PE2(config)#interface Loopback1
PE2(config-if-loopback1)#ip address 200.200.1.1 255.255.255.255
PE2(config-if-loopback1)#exit
PE2(config)#router ospf
PE2(config-router)#network 200.200.1.1/32 area 0
PE2(config-router)#network 202.200.1.0/24 area 0
PE2(config-router)#exit
PE2(config)#router bgp 100
PE2(config-router)#address-family vpnv4 unicast
PE2(config-router-af)#neighbor 200.200.1.1 activate
PE2(config-router-af)#exit-address-family
PE2(config-router)#address-family ipv4 vrf VRF-A
PE2(config-router-af)#neighbor 192.168.101.2 remote-as 60101
PE2(config-router-af)#no neighbor 192.168.101.2 send-community extended
PE2(config-router-af)#exit-address-family
PE2(config-router)#exit
PE2(config)#router ldp
```

The configuration of CE2 is as follows:

```
CE2#config
CE2(config)#interface vlan 1
CE2(config-if-Vlan1)#ip address 192.168.101.2 255.255.255.0
CE2(config-if-Vlan1)#exit
CE2(config)#interface Loopback1
CE2(config-if-Loopback1)#ip address 1.1.1.1 255.255.255.255
CE2(config-if-Loopback1)#exit
CE2(config)#router bgp 60101
CE2(config-router)#network 10.1.3.0/24
CE2(config-router)#neighbor 192.168.101.1 remote-as 100
```

The configuration of IGW is as follows:

```
IGW#config
IGW(config)#interface Vlan1
IGW(config-if-Vlan1)#ip address 202.200.3.2 255.255.255.0
IGW(config-if-Vlan1)#exit
IGW(config)#interface Vlan2
IGW(config-if-Vlan2#ip address 150.1.1.1 255.255.255.0
IGW(config-if-Vlan2#exit
IGW(config)#router ospf
```

IGW(config-router)#network 202.200.3.0 0.0.0.255 area 0
IGW(config-router)#exit
IGW(config)#router bgp 100
IGW(config-router)#neighbor 202.200.2.2 remote-as 100
IGW(config-router)#neighbor 202.200.2.2 default-originate

10.4.3.2 VRF Internet Access Mode 1

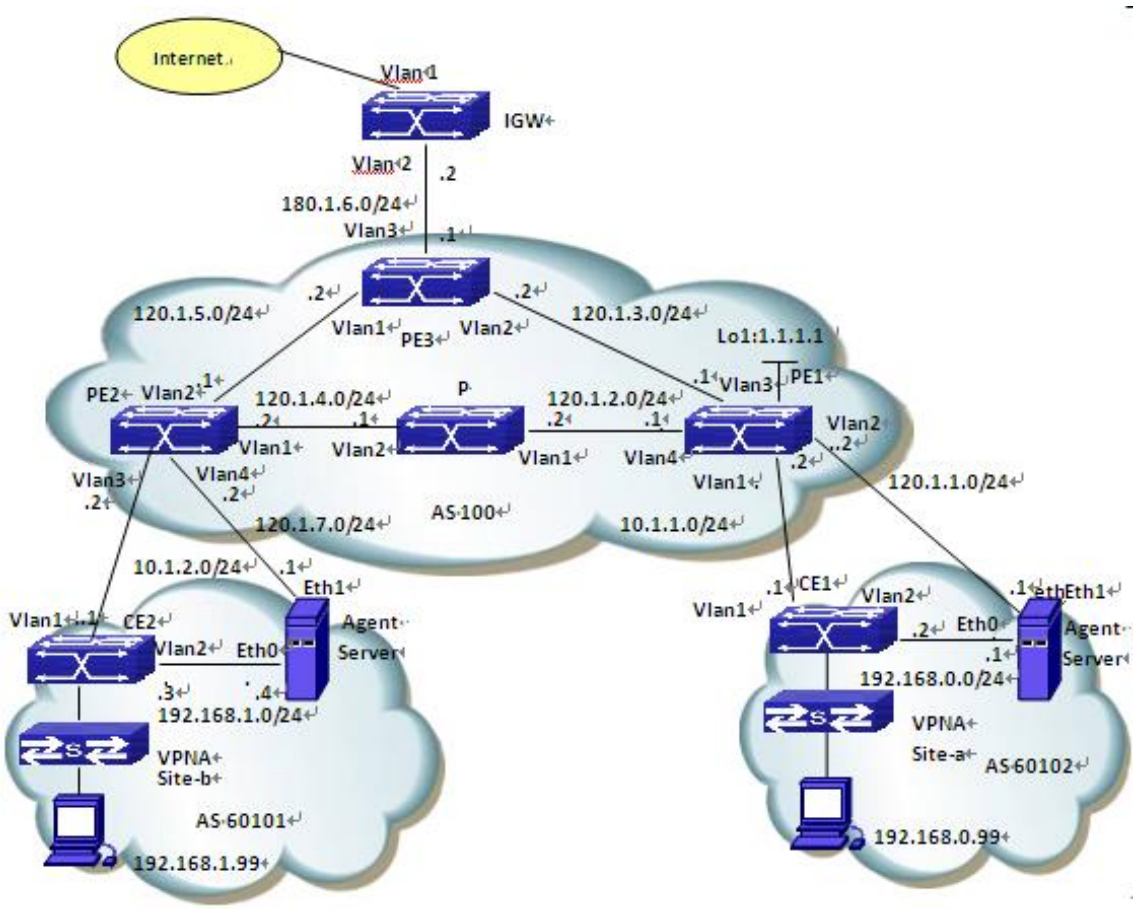


Fig 10-24 VRF Internet Access Mode 1

Site-a and site-b belong to VPNA; their users can intercommunicate and all need to access the Internet. Configure proxy servers separately in site-a and site-b to realize NAT when their users access Internet with the private network addresses.

The configuration of CE1 is as follows:

CE1#config
CE1(config)#interface Vlan1
CE1(config-if-Vlan1)#ip address 10.1.1.1 255.255.255.0
CE1(config-if-Vlan1)#exit
CE1(config)#interface Vlan2
CE1(config-if-Vlan2)#ip address 192.168.0.2 255.255.255.0
CE1(config-if-Vlan2)#exit
CE1(config)#interface loopback1

```
CE1(config-if-Loopback1)#ip address 11.11.11.11 255.255.255.255
CE1(config-if-Loopback1)#exit
CE1(config)#ip route 0.0.0.0/0 192.168.0.1
CE1(config)#router bgp 60101
CE1(config-router)#neighbor 10.1.1.2 remote-as 100
CE1(config-router)#network 192.168.0.0/24
```

The configuration of PE1 is as follows:

```
PE1#config
PE1(config)#ip vrf VPNA
PE1(config-vrf)#rd 100:10
PE1(config-vrf)#route-target both 100:10
PE1(config-vrf)#exit
PE1(config)#interface Vlan1
PE1(config-if-Vlan1)#ip vrf forwarding VPNA
PE1(config-if-Vlan1)#ip address 10.1.1.1 255.255.255.0
PE1(config-if-Vlan1)#exit
PE1(config)#interface Vlan2
PE1(config-if-Vlan2)#ip vrf forwarding VPNA fallback global
PE1(config-if-Vlan2)#ip address 120.1.1.2 255.255.255.
PE1(config-if-Vlan2)#exit
PE1(config)#interface Vlan3
PE1(config-if-Vlan3)#ip address 120.1.3.1 255.255.255.0
PE1(config-if-Vlan2)#exit
PE1(config)#interface Vlan4
PE1(config-if-Vlan4)#label-switching
PE1(config-if-Vlan4)#ldp enable
PE1(config-if-Vlan4)#ip address 202.200.1.2 255.255.255.0
PE1(config-if-Vlan4)#exit
PE1(config)#interface loopback1
PE1(config-if-Loopback1)#ip address 1.1.1.1 255.255.255.255
PE1(config-if-Loopback1)#exit
PE1(config)#router ospf
PE1(config-router)#redistribute static
PE1(config-router)#network 1.1.1.1/32 area 0
PE1(config-router)#network 120.1.2.0/24 area 0
PE1(config-router)#network 120.1.3.0/24 area 0
PE1(config-router)#exit
PE1(config)#router bgp 100
PE1(config-router)#neighbor 2.2.2.2 remote-as 100
PE1(config-router)#neighbor 2.2.2.2 update-source 1.1.1.1
PE1(config-router)#address-family vpnv4 unicast
PE1(config-router-af)#neighbor 2.2.2.2 activate
PE1(config-router-af)#exit-address-family
PE1(config-router)#address-family ipv4 vrf VPNA
```

```
PE1(config-router-af)#network 120.1.1.0/24
PE1(config-router-af)#neighbor 10.1.1.1 remote-as 60101
PE1(config-router-af)#no neighbor 10.1.1.1 send-community extended
PE1(config-router-af)#exit-address-family
PE1(config-router)#exit
PE1(config)#router ldp
PE1(config-router)#exit
PE1(config)#ip route 120.1.1.0/24 vlan 2 120.1.1.1
```

The configuration of P is as follows:

```
P#config
P(config)#interface Vlan1
P(config-if-Vlan1)#label-switching
P(config-if-Vlan1)#ldp enable
P(config-if-Vlan1)#ip address 120.1.2.2 255.255.255.0
P(config-if-Vlan1)#exit
P(config)#interface Vlan2
P(config-if-Vlan2)#label-switching
P(config-if-Vlan2)#ldp enable
P(config-if-Vlan2)#ip address 120.1.4.1 255.255.255.0
P(config-if-Vlan2)#exit
P(config)#router ospf
P(config-router)#network 0.0.0.0/0 area 0
P(config-router)#exit
P(config)#router ldp
```

The configuration of PE2 is as follows:

```
PE2#config
PE2(config)#ip vrf VPNA
PE2(config-vrf)#rd 100:10
PE2(config-vrf)#route-target both 100:10
PE2(config-vrf)#exit
PE2(config)#interface Vlan1
PE2(config-if-Vlan1)#label-switching
PE2(config-if-Vlan1)#ldp enable
PE2(config-if-Vlan1)#ip address 120.1.4.2 255.255.255.0
PE2(config-if-Vlan1)#exit
PE2(config)#interface Vlan2
PE2(config-if-Vlan2)#ip address 120.1.5.1 255.255.255.0
PE2(config-if-Vlan2)#exit
PE2(config)#interface Vlan3
PE2(config-if-Vlan3)#ip vrf forwarding VPNA
PE2(config-if-Vlan1)#ip address 10.1.2.2 255.255.255.0
PE2(config-if-Vlan1)#exit
PE2(config)#interface Vlan4
PE2(config-if-Vlan4)#ip vrf forwarding VPNA fallback global
```

```
PE2(config-if-Vlan4)#ip address 120.1.7.2 255.255.255.0
PE2(config-if-Vlan4)#exit
PE2(config)#interface Loopback1
PE2(config-if-Loopback1)#ip address 2.2.2.2 255.255.255.255
PE2(config-if-Loopback)#exit
PE2(config)#router ospf
PE2(config-router)#redistribute static
PE2(config-router)#network 2.2.2.2/32 area 0
PE2(config-router)#network 120.1.4.0/24 area 0
PE2(config-router)#network 120.1.5.0/24 area 0
PE2(config-router)#exit
PE2(config)#router bgp 100
PE2(config-router)#neighbor 1.1.1.1 remote-as 100
PE2(config-router)#neighbor 1.1.1.1 update-source 2.2.2.2
PE2(config-router)#address-family vpnv4 unicast
PE2(config-router-af)#neighbor 1.1.1.1 activate
PE2(config-router-af)#exit-address-family
PE2(config-router)#address-family ipv4 vrf VPNA
PE2(config-router-af)#network 120.1.7.0/24
PE2(config-router-af)#neighbor 10.1.2.1 remote-as 60102
PE2(config-router-af)#no neighbor 10.2.1.1 send-community extended
PE2(config-router-af)#exit-address-family
PE2(config-router)#exit
PE2(config)#router ldp
PE2(config-router)#exit
PE2(config)#ip route 120.1.7.0/24 vlan 4 120.1.7.1
```

The configuration of PE3 is as follows:

```
PE3#config
PE3(config)#interface Loopback1
PE3(config-if-Loopback1)#ip address 3.3.3.3 255.255.255.255
PE3(config-if-Loopback1)#exit
PE3(config-if-Vlan1)#interface Vlan1
PE3(config-if-Vlan1)#ip address 120.1.5.2 255.255.255.0
PE3(config-if-Vlan1)#exit
PE3(config)#interface Vlan2
PE3(config-if-Vlan2)#ip address 120.1.3.2 255.255.255.0
PE3(config-if-Vlan2)#exit
PE3(config)#interface Vlan3
PE3(config-if-Vlan3)#ip address 180.1.6.1 255.255.255.0
PE3(config-if-Vlan3)#exit
PE3(config)#router ospf 1
PE3(config-router)#default-information originate
PE3(config-router)# network 0.0.0.0/0 area 0
PE3(config-router)#exit
```

```
PE3(config)#router bgp 100
PE3(config-router)#network 120.1.1.0 mask 255.255.255.0
PE3(config-router)#network 120.1.7.0 mask 255.255.255.0
PE3(config-router)#neighbor 180.1.6.2 remote-as 200
PE3(config-router)#exit
PE3(config)#ip route 0.0.0.0/0 180.1.6.2
```

The configuration of CE2 is as follows:

```
CE2#config
CE2(config)#interface Vlan1
CE2(config-if-Vlan1)#ip address 10.1.2.1 255.255.255.0
CE2(config-if-Vlan1)#exit
CE2(config)#interface Vlan2
CE2(config-if-Vlan2)#ip address 192.168.1.3 255.255.255.0
CE2(config-if-Vlan2)#exit
CE2(config-if-Loopback1)#interface Loopback1
CE2(config-if- Loopback1)#ip address 22.22.22.22 255.255.255.255
CE2(config-if- Loopback1)#exit
CE2(config)#ip route 0.0.0.0/0 192.168.1.4
CE2(config)#router bgp 60101
CE2(config-router)#neighbor 10.1.2.2 remote-as 100
CE2(config-router)#network 192.168.1.0/24
CE2(config-router)#exit
```

The configuration of IGW is as follows:

```
IGW#config
IGW(config)#interface Vlan1
IGW(config-if-Vlan1)#ip address 180.1.5.2 255.255.255.0
IGW(config-if-Vlan1)#exit
IGW(config)#interface Vlan2
IGW(config-if-Vlan2)#ip address 180.1.6.2 255.255.255.0
IGW(config-if-Vlan2)#exit
IGW(config)#router bgp 200
IGW(config-router)#neighbor 180.1.6.1 remote-as 100
IGW(config-router)#exit
```

10.4.4 Public Network Access Troubleshooting

When configuring and using Public Network Access, some problems like incorrect physical connections, configuration errors may cause it to fail, so please pay attention to the following notices to avoid them:

- ☞ First, make sure the regular MPLS BGP VPN works correctly, and the intercommunication is normal in the private network. If the communication in VPN fails, please refer to the help on MPLS BGP VPN troubleshooting.
- ☞ Second, check the public network access mode in use is non-VRF or VRF, for

their configurations differ a lot.

- ☞ In non-VRF mode, please remember to configure filter policy on the non-VRF interface of PE-CE, to block the private network route and traffic from entering PE through the public network interface. Otherwise, there might be security threats. Besides, make sure the advertisement of default routes and the NAT configuration to IGW are correct.
- ☞ In VRF mode, please make sure to use “ip vrf forwarding vrf_name fallback global” command while configuring the private network interface, to prevent look up the global route table for a second time if the attempt to find the private network route fails. Besides, make sure the advertisement of default routes and the NAT configuration to IGW are correct.
- ☞ At last, if all above steps are correct, CE will be able to access Internet. No matter which networking mode mentioned above is used, other CE access Internet after forwarding traffic to PE via VPN; the traffic from Internet should also be forwarded after passing PE.

10.5 VPLS

10.5.1 VPLS Overview

Nowadays IP network has been all over the world. How to use the existing IP network to offer a low-cost private network that becomes a focuss to providers. As a result, MPLS VPN—a technology which can offer VPN service in IP network, easy to set rate and configure simply. There are two types in this technology: MPLS L3VPN and MPLS L2VPN. Since MPLS L3VPN intervenes users’ internal route management, provider’s management is complex. MPLS L2VPN in a traditional way of VLL, offer a point-to-point L2VPN service, it can make the connection between two stations liking the direct link connection. However, it cannot offer the switchover between multi-points for providers. VPLS develops on the basis of the traditional MPLS L2VPN, it can implement VPN network (multi-points to multi-points). So VPLS provides a better solution for providers.

VPLS, namely Virtual Private LAN Service, is a service to offer LAN in MPLS network, accurately, it’s a L2VPN technology based on Ethernet. VPLS technology includes the advantages of Ethernet and MPLS, and it can make user’s network communicate with each other at different location, as they are connected with each other directly. Furthermore, VPLS enables users to extend their LAN to MAN, or even WAN.

VPWS is a special example of VPLS, conversely, VPLS is extension of VPWS. Since VPLS offers a VPN group network (multi-points to multi-points) while VPWS only offers a point-to-point L2VPN solution, in this way ,VPLS networks can be looked as a switch to user CE, but VPWS

correspond as an Ethernet cable.

VPLS chooses LDP signaling and BGP signaling to form PW. A signaling based on LPD protocol establishes virtual circuit by establishing a point-to-point LDP session between two PEs, since LDP protocol is set earlier and application is simple, many products support this protocol. In terms of compatibility, we adopt a signaling mechanism based on LDP.

10.5.1.1 Basic Concept of VPLS

- **VPLS（Virtual Private LAN Service）**

VPLS is a service to offer LAN in MPLS network, it can make users to access network from dispersed points at the same time, as they are connected with each other directly. Furthermore, VPLS enables users to extend their LAN to MAN, or even WAN.

- **VC（Virtual Circuit）**

It is a unilateralism logic connection between two points, a pair of reverse VC forms a PW.

- **S-TAG(Service Tag)、S-VID(Service VLAN ID)**

They are offered by provider's network to identify users.

- **P-TAG(Provider Tag)、P-VID(Provider VLAN ID)**

They are expected VLAN Tags for the other peer.

- **VFI（Virtual Forward Instance）**

Each VFI offers separated VPLS service, and it fulfills the function of Ethernet bridging. Using VFI can map VPLS's actual access links to each PW, in turn, it can push PW to the end and map PW to the actual access link.

- **UPE（User facing-Provider Edge）**

It is an aggregation device for accessing VPN.

- **NPE（Network Provider Edge）**

It is a core PE device at the core field edge of VPLS network, offer the transparent VPLS transmission between core networks.

- **MTU（Maximum Transmission Unit）**

Maximum transmission unit.

- **QinQ（802.1Q in 802.1Q）**

It is a tunnel protocol based on 802.1Q encapsulation, it is able to offer point to multi-point service for L2VPN. It encapsulates user's VLAN Tag in private network to public network, then the final packets with two tags go through provider's backbone network, which provides users with a simpler layer 2 VPN tunnel.

- **PW Signaling**

It is used to establish and maintain PW, and is the basis of VPLS. It can be used to discover VFI's peer PE device automatically. There are LDP and BGP in PW signaling so far. This work supports LDP only.

- **VPWS (Virtual Private Wire Service)**

It is a point-to-point service for L2VPN. When one peer (AC or PW) receives packets, it will be forwarded directly to the other peer (PW or AC). Compared with VPLS, VPWS does not require to learn and look for MAC address, furthermore, there is no broadcast, multicast, etc., so it is

more efficient.

10.5.1.2 Basic Network Model of VPLS

The network model of VPLS has five parts: CE、PE、P、AC and PW. The following is the basic concept.

- **CE（Custom Edge）**
  It is an edge device connected directly with providers. It can be a router, a switch or a host.
- **PE（Provider Edge）**
  It is an edge device offered by providers, connected with CE, and responsible for accessing VPN. It can map and forward packets from public tunnel to private tunnel.
- **P（Provider）**
  It a backbone router in provider's network, it does not connect with CE. It only needs to have basic MPLS forwarding capability and does not maintain VPN information.
- **AC（Attachment Circuit）**
  In L2VPN, CE is able to access to PE via AC. AC is used to transport frames between CE and PE, and it can be a physic or logical link.
- **PW（Pseudo Wire）**
  Simply, PW means VC and tunnel, which can be LSP, GRE or CR-LSP. For VPLS, PW is like a direct tunnel between ACs to complete layer 2 data passthrough of users.

There are two models in rfc4762: one is fully connection to PW and the other is layered VPLS model. As shown in figure 1 and 2

10.5.1.3 Advantages of VPLS

VPLS advantages:

- VPLS uses the Ethernet interface to users, simplifying LAN/WAN boundary, can support a fast and flexible service deployment.
- VPLS gives the right of controlling and maintaining route policy to users, simplifying the network management from providers.
- All CEs included in VPLS service are part of a sub-network, simplifying IP address configuration.
- VPLS service does not participate in IP addressing and route.

10.5.1.4 Basic VPLS Network Model of Fully PE Connection

All PEs connect with each other logically, they can learn MAC addresses and forward packets among multi-points. MPLS network offers tunnel to pass through packets between VPN stations. P devices are similar to those in L3VPN, they are responsible for forwarding MPLS packets but do not participate in learning or exchanging MAC addresses. In order to overlap MAC addresses in VPNs, the forwarding tables are independent.

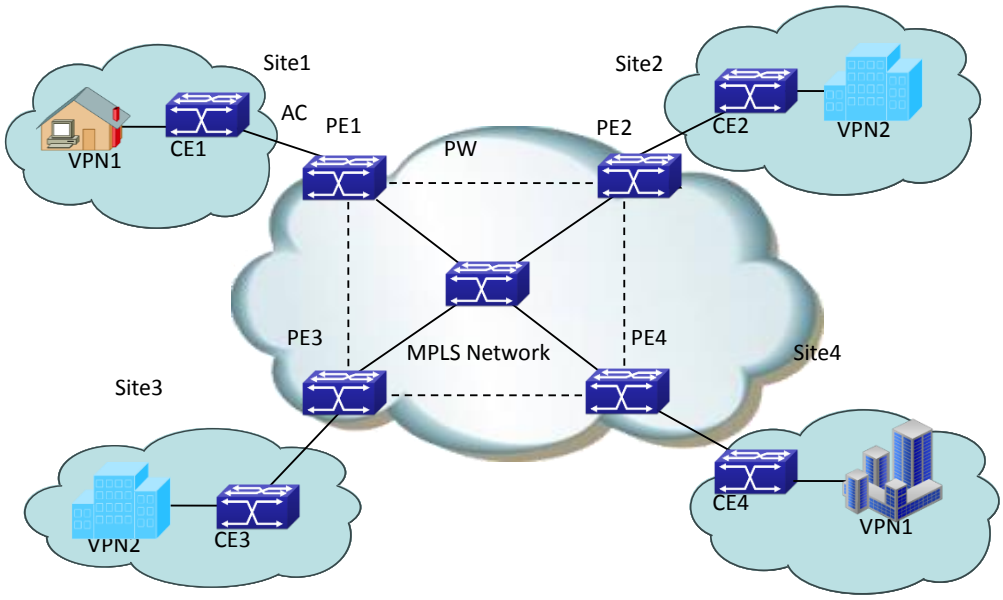


Figure 10-25 VPLS model of fully connection PE

Fully connection VPLS model avoids loopback through level division forwarding. If a PE receives packet from a PW, then this packet will not be forwarded to other PWs relevant to VFI. That is to say, any two PEs communicate with each other via a PW directly, do not forward packet via the third PE. That is why VFIs require fully connection PW.

10.5.1.5 Layered VPLS Model

All NPEs connect with each other logically. UPE establishes a virtual connection with closest NPE, and exchange packets with VPN station through NPE, so network topology is getting stratificational. UPE is mostly used to access VPN and has lower performance requirement, while NPE is mostly used to aggregate flow, and has higher performance requirement. Additionally, to ensure a strong network, we can add a link backup between UPE and NPE. The virtual connection between UPE and NPE can be established according to QinQ or LDP.

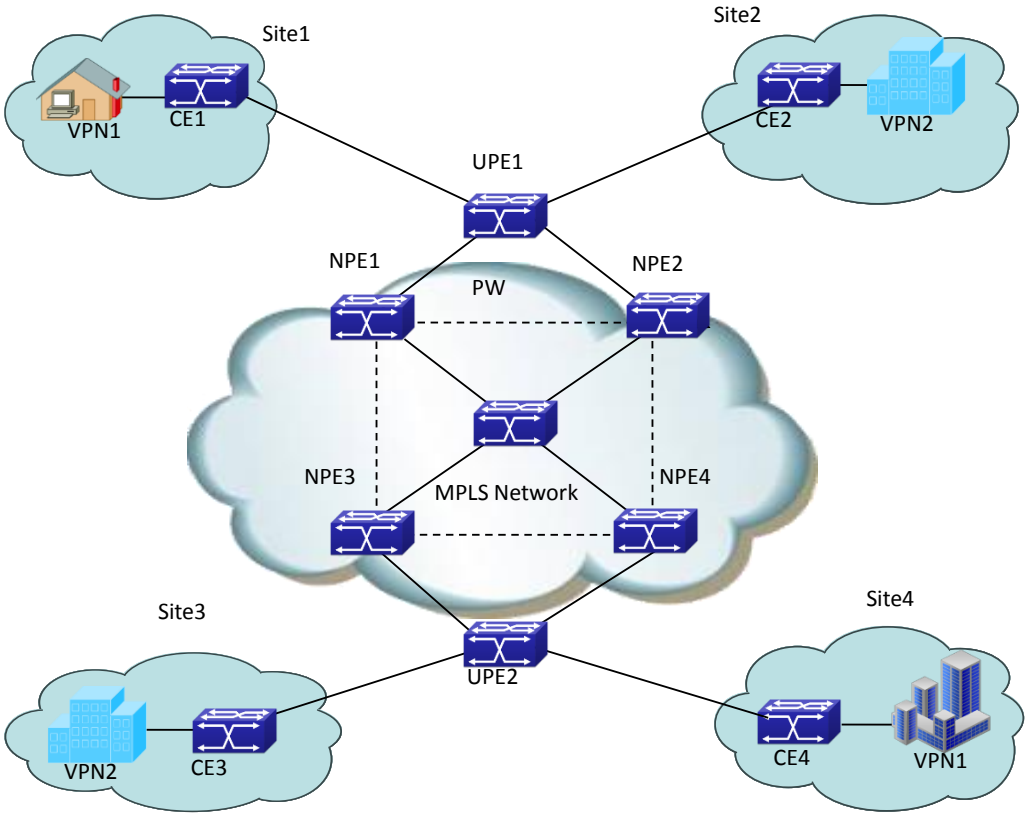


Figure 10-206 Layered VPLS model

According to the different connections between UPE and NPE, H-VPLS is divided to LSP and QinQ access method.

☞ **LSP access method**

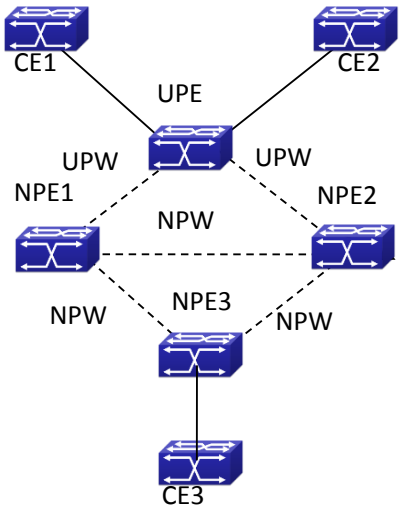


Figure 10-217 LSP access method

As shown in figure, UPE works as an aggregation device, it connects NPE1 and NPE2 via LSP. UPE must establish virtual U-PW with NPE1 and NPE2 separately (U-PW connection needs the new VFI on UPE and NPE to set peers, and PWIDs on two devices must be the same), UPE does

not establish the virtual connection with other devices.

The process of forwarding data for LSP access mode is as follows:

(1) UPE sends packet coming from CE to NPE1, and marks the corresponding VC tag (it is assigned by NPE1, and work as an unattached tag for multiplex PW) related to U-PW.

(2) When NPE1 receives packet, it will analyze VC tag and decide which VFI it belongs to, then push it in the corresponding VC tag according to destination MAC, finally, forward packet.

(3) After NPE1 receives packet from N-PW, it marks the corresponding VC tag, then forward packet to UPE, finally to CE.

When data exchanging between CE1 and CE2 becomes the exchanging between the local CEs, since UPE has a bridging function, it can complete packet forwarding between CEs. However, for unknown data packets or broadcast packets of destination MAC, UPE still forward them to NPE1 via U-PW when packets are broadcasted to CE2 via bridge. Finally NPE1 will copy and forward packets to CEs on the other peer.

☞ **QinQ access method**

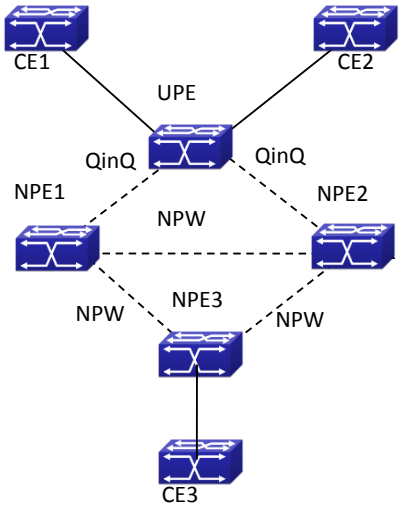


Figure 10-228 QinQ access method

As shown in figure, UPE is a standard bridging device, it establishes point-to-point Ethernet QinQ connection between UPE and PEs (that is, apply QinQ to CE interface while apply VLAN access mode to PE1). Packets received by UPE will be marked an outer VLAN tag, when they are forwarded to PE1, the outer VLAN tag can be described to a VLAN tag of providers according to VLAN access mode, namely, a service delimiter, according to the delimiter, packets are mapped to relevant VFI, then this VFI makes up its mind how to forward packets (unicast or multicast).

The process of forwarding data for QinQ is as follows:

(1) Enable QinQ at CE access port, add the unattached multiplex tag for packets, and pass through them to PE1 via QinQ tunnel between UPE and PE1.

(2) When PE1 receives packet, it will analyze the VLAN tag and decide which VFI it belongs to, then push relevant PW tag in according to destination MAC address, finally, it will be forwarded.

(3) After PE1 receives packet from PW, it will decide which VFI it belongs to according to PW tag, and marks a VLAN tag according to destination MAC, then forward packet via QinQ tunnel to UPE, finally to CE.

When data exchanging between CE1 and CE2 becomes the exchanging between the local

CEs, since UPE has a bridging function, it can complete packet forwarding between CEs. However, for unknown data packets or broadcast packets of destination MAC, UPE still forward them to PE1 via QinQ when packets are broadcasted to CE2 via bridge. Finally PE1 will copy and forward packets to CEs on the other peer.

☞ **Method of avoidance loopback for H-VPLS model**

Compared to full connection, method of avoidance loopback for the layered VPLS model is a different. Since H-VPLS only needs to establish a full connection between NPEs, do not establish that between UPE and NPE, packets received from PW are not forwarded to relevant PW connected with other NPE, however, it allows to forward packets to relevant PW connected with UPE. Furthermore, packets received from PW connecting to UPE can be forwarded to PW connected with other NPEs.

10.5.1.6 Packet Forwarding of VPLS

In VPLS model, there are two packets encapsulation methods in AC and PW.
In AC there are VLAN and Ethernet access methods. Introduction is as follows:

☞ **VLAN access**

Ethernet frame header with a VLAN TAG sent by CE to PE or PE to CE, it is a delimiter added by provider's device to distinguish users, and we call it as S-TAG.

☞ **Ethernet access**

Ethernet frame header without a delimiter sent by CE to PE or PE to CE. If there is a VLAN TAG in it, however, this TAG is an inner one, and it is meaningless to PE. We call it as U-TAG.

There are two packets encapsulation methods in PW, they are Raw and Tagged modes. Introduction is as follows:

☞ **Raw mode**

A frame transported by PW cannot without S-TAG: for packets of CE, if packets with delimiter are received, they will remove the delimiters and push PW and tunnel tags in before forwarding. If packets without delimiter are received, they will push PW and tunnel tags in directly before forwarding. For downlink packets of PE, they can add delimiters or not to forward to CE according to the actual configuration, but they cannot rewrite or remove the existent tags.

☞ **Tagged mode**

A frame transported in PW has to own an S-TAG: for packets of CE, if packets with delimiter are received, it will be compared with the expected VLAN TAG from the peer. If they are same, keep S-TAG, else rewrite S-TAG as the expected VLAN TAG. If the other peer does not transfer the expected VLAN TAG, it will fill in null TAG (value of TAG is 0). Finally push PW and tunnel tags in before forwarding. On the condition that there is not expected VLAN TAG, TAG is null (TAG equals 0). If packets without delimiter are received, add an expected VLAN TAG or a null TAG, and then push PW and tunnel tags in before forwarding. For downlink packets of PE, they can rewrite, remove or keep delimiters according to the actual configuration before forwarding.

According to the different of AC access mode and packet encapsulation mode in PW, VPLS packet forwarding mode can be divided into the following four types.

☞ **Forward packets in Ethernet access Raw mode**

As shown in figure, AC uses Ethernet access mode, and packet encapsulation mode is Raw mode on PW, the process of packet forwarding from CE1 to CE4 is described as follows:

1. CE1 sends packets with U-Tag to PE1
2. PE1 chooses an appropriate PW according to destination MAC address and adds a VC tag in packets.
3. In order to forward packets in public network via MPLS tunnel, PE1 adds a public network tunnel tag and forwards packets to PE4.
4. After PE4 receives packets, it finds out the relevant VFI according to the VC tag, and sends packets with U-TAG to CE4.

☞ **Forward packets in Ethernet access Tagged mode**

As shown in figure, AC uses Ethernet access mode. When packet encapsulation mode is Tagged mode on PW, the process is similar to packets forwarded in Ethernet access mode and Raw mode. The difference is that frame in PW has to be with S-TAG. After PE1 receives packets without S-TAG, first, it will add an expected VLAN TAG or a null TAG, and then, it will push two layers MPLS tag before forwarding. When PE4 receives packets, it will remove the two layers tag and S-TAG before forwarding them to CE4.

☞ **Forward packets in VLAN access Raw mode**

As shown in figure, AC uses VLAN access mode, messages packet encapsulation mode is RAW mode, the process of packets forwarded from CE1 to CE4 is described as follows:

1. Packets sent by CE1 have to be with S-TAG. If packets forwarded by CE1 without VLAN TAG or with unmatched VLAN TAG, it executes the normal layer 2 forwarding process.
2. After PE1 receives packets, it removes S-TAG and adds a two layer MPLS tags before forwarding them to PE4 via public network MPLS tunnel.
3. After PE4 removes two layer MPLS tags in packets, adds S-TAG before forwarding them to CE4.

☞ **Forward packets in VLAN access Tagged mode**

As shown in figure, AC uses VLAN access mode. When packet encapsulation mode is Tagged mode on PW, the process is similar to packets forwarded in VLAN access mode and Raw mode. The difference is that frame in PW with S-TAG. After PE1 receives packets, it will be compared with the expected VLAN TAG sent by the peer PE, if they are same, keep S-TAG, else rewrite S-TAG as the expected VLAN TAG. On the condition that there is not expected VLAN TAG, it will fill in a null TAG (value of TAG is 0).

10.5.2 VPLS Configuration

VPLS configuration task sequence:

1. Globally enable route protocol (required)
2. Basic configuration of MPLS (required)
   (1) Enable MPLS globally
   (2) Enable tag exchanging on interface
3. Configure LDP session (required)
   (1) Enter ldp view in configuration mode, which stands for enabling LDP function
   (2) Configure remote LDP neighbor and enable targeted-peer command

(3) Enable LDP command in layer 3 interface view

4. Create PW template (optional)

(1) Create pw-class and enter pw view

(2) Set transmission mode for pw-class

5. Configure VFI (required)

(1) Create VFI, specify VFI ID and enter VFI view

(2) Configure the end point for remote PW (usually it is same as targeted peer of LDP)

(3) Specify pw-class for each peer

(4) Set PW encapsulation mode for VFI

6. Configure users' access mode and bind VFI (VPLS required)

(1) Enter port view

(2) Configure binding VFI for port

7. Configure binding PW for port (VPWS required）

(1) Create L2VC and configure VPWS's peer PE

(2) Enter port view

(3) Bind port with the specified PW and enter access mode

1. Globally enable route protocol (required)

Command	Description
Global configuration mode	
router ospf	Enter routing configuration mode for OSPF
OSPF routing configuration mode	
network 0.0.0.0/0 area 0	Configure the interface segment address enabled OSPF, by default, OSPF is disabled on interface.

2. Global MPLS (required)

Command	Description
Global configuration mode	
mpls enable **no mpls enable**	Enable MPLS protocol, no command disables MPLS protocol.
Layer 3 interface view	
label-switching enable **no label-switching**	Enable tags exchanging function, no command disables the function by default.

3. Configure LDP session (required)

(1) Enter ldp view in configuration mode, which stands for enabling LDP function

(2) Configure remote LDP neighbor and enable targeted-peer command

(3) Enable LDP command in layer 3 interface view

Command	Description
Global configuration mode	
router ldp **no router ldp**	Enable LDP protocol, no command disables LDP.
targeted-peer 1.1.1.1	Configure remote LDP neighbor. By default, there is no remote LDP neighbor.

Interface configuration mode	
ldp {enable | disable}	Enable LDP protocol on interface, no command disables LDP protocol.

4. Create PW template (optional)

(1) Create pw-class and enter pw view

(2) Set transmission mode for pw-class

Command	Description
Global configuration mode	
pw-class <pw-class-name> **no pw-class <pw-class-name>**	Create pw-class. By default, there is no PW template.
PW template configuration mode	
transport-mode {ethernet | vlan}	Configure packet encapsulation mode on PW template, ethernet corresponds to raw mode and vlan corresponds to tagged mode.

5. Configure VFI (required)

(1) Create VFI, specify VFI ID and enter VFI view

(2) Configure the end point for remote PW (usually it is same as targeted peer of LDP)

(3) Specify pw-class for each peer

(4) Set PW encapsulation mode for VFI

Command	Description
Global configuration mode	
vfi <vfi-name> <vfi-id> **no vfi <vfi-name>**	Create VFI and specify VFI ID. There is no VFI.
VFI configuration mode	
peer ip-address [pw-id pw-id] [no-split-horizon] [pw-class class-name]	Peer PE included in VPLS instance can configure whether enable level division (enabled by default) and pw template.
transport-mode {ethernet | vlan}	Configure packet encapsulation mode in PW.

6. Configure users' access mode and bind VFI (VPLS required)

(1) Enter port view

(2) Configure binding VFI for port

Command	Description
Port configuration mode	
xconnect vfi vfi-id [mode {ethernet | vlan [svid svid]}]	Configure a port to bind VFI and configure AC access mode. By default, port does not bind any VFI.

7. Configure binding PW for port (VPWS required）

(1) Create L2VC and configure VPWS's peer PE

(2) Enter port view

(3) Bind port with the specified PW and enter access mode

Command	Description
Global configuration mode	
l2-vc ip-address pw-id pw-id [group group-id] [pw-class class-name]	Create L2VC and configure VPWS's peer PE. By default, there is no configuration for peer PE.
Enter port view	
xconnect l2-vc pw-id <pw-id> [mode {ethernet | vlan [svid <svid>]}]	Bind port with the specified PW and enter access mode. By default, port is not bound to any PW.

10.5.3 Typical Examples of VPLS

10.5.3.1 Full Connection for VPLS Configuration

10.5.3.1.1 Network Requirement

- CE1 and CE2 belong to two different stations, while they belong to the same VPN1
- CE1 and CE2 can be accessed to PE via port Eth1/0/1
- CE1 can access to PE1 in Ethernet mode, while CE2 can access to PE2 in VLAN mode, Svid equals 200.
- Packet encapsulation mode between PE1 and PE2 is RAW (Ethernet) mode, while that between PE3 and PE1 (PE2) is Tagged (Vlan) mode.
- It requires that establish a layer 2 VPN1 by configuring VPLS, making CE1 and CE2 access each other in layer 2 mode.

10.5.3.1.2 Group Network Diagram

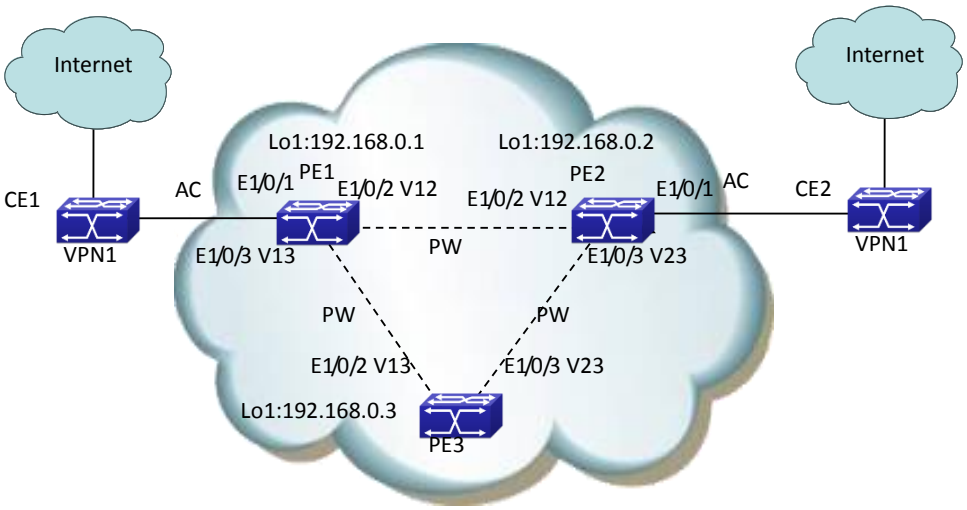


Figure 10-239 full connection VPLS model for PE

10.5.3.1.3 Configuration Steps

Please configure IP address and mask for interfaces including VLAN and Loopback according to the above figure. The specific configuration is omitted. This is for VPLS examples in LDP.

(1) PE1 configuration

#Create PW template

```
PE1(config)#pw-class c1
PE1(config-class) #transport-mode ethernet
PE1(config-class)#exit
PE1(config)#pw-class c2
PE1(config-class)#transport-mode vlan
PE1(config-class)#exit
```

#Configure the routing

```
PE1(config)#router ospf
PE1(config-router)#router-id 192.168.0.1
PE1(config-router)#network 0.0.0.0/0 area 0
PE1(config-router)#exit
```

#Configure MPLS capability and label switching capability (Interface connected with CE enables label switching capability in non-H-VPLS)

```
PE1(config)#mpls enable
PE1(config)#int vlan 12
PE1(config-if-vlan12)#label-switching
PE1(config-if-vlan12)#exit
PE1(config)#int vlan 13
PE1(config-if-vlan13)#label-switching
PE1(config-if-vlan13)#exit
```

#Configure LDP

```
PE1(config)#router ldp
PE1(config-router)#router-id 192.168.0.1
PE1(config-router)#targeted-peer 192.168.0.2
PE1(config-router)#targeted-peer 192.168.0.3
PE1(config-router)#exit
PE1(config)#int vlan 12
PE1(config-if-vlan12)#ldp enable
PE1(config-if-vlan12)#exit
PE1(config)#int vlan 13
PE1(config-if-vlan13)#ldp enable
PE1(config-if-vlan13)#exit
```

#Configue instance V1 and remote PE in LDP

```
PE1(config)#vfi v1 100
PE1(config-vfi)#peer 192.168.0.2 pw-class c1
PE1(config-vfi)#peer 192.168.0.3 pw-class c2
```

```
PE1(config-vfi)#exit

#Bind VPLS instance mode as ethernet
PE1(config-if-ethernet1/0/1)#xconnect vfi 100 mode ethernet
```

(2) PE2 configuration

```
#Create PW template
PE2(config)#pw-class c1
PE2(config-class)#transport-mode ethernet
PE2(config-class)#exit
PE2(config)#pw-class c2
PE2(config-class)#transport-mode vlan
PE2(config-class)#exit

#Configure the routing
PE2(config)#router ospf
PE2(config-router)#router-id 192.168.0.2
PE2(config-router)#network 0.0.0.0/0 area 0
PE2(config-router)#exit
```

#Configure MPLS capability and label switching capability (Interface connected with CE does not enable label switching capability in non-H-VPLS)

```
PE2(config)#mpls enable
PE2(config)#int vlan 12
PE2(config-if-vlan12)#label-switching
PE2(config-if-vlan12)#exit
PE2(config)#int vlan 23
PE2(config-if-vlan23)#label-switching
PE2(config-if-vlan23)#exit
```

#Configure LDP

```
PE2(config)#router ldp
PE2(config-router)#router-id 192.168.0.2
PE2(config-router)#targeted-peer 192.168.0.1
PE2(config-router)#targeted-peer 192.168.0.3
PE2(config-router)#exit
PE2(config)#int vlan 12
PE2(config-if-vlan12)#ldp enable
PE2(config-if-vlan12)#exit
PE2(config)#int vlan 23
PE2(config-if-vlan23)#ldp enable
PE2(config-if-vlan23)#exit
```

#Configue instance V1 and remote PE in LDP

```
PE2(config)#vfi v1 100
PE2(config-vfi)#peer 192.168.0.1 pw-class c1
PE2(config-vfi)#peer 192.168.0.3 pw-class c2
PE2(config-vfi)#exit

#Bind VPLS instance mode as VLAN on port, Svid is 200
PE2(config-if-etherne1/0/1)#xconnect vfi 100 mode vlan svid 200
```

(3) PE3 configuration

It is similar to PE1 and PE2.

10.5.3.2 Access H-VPLS with LSP

10.5.3.2.1 Network Requirement

- CE1 and CE2 belong to two stations, while they belong to the same VPN1
- CE1 and CE2 can access to PE via port Eth1/0/1
- N-PE1,N-PE2 and N-PE3 create a full connection VPLS network
- U-PE is an access device of user, it accesses N-PE1 with layered VPLS PW mode
- CE1 can access to U-PE in Ethernet mode, while CE2 can access to N-PE3 in VLAN mode, Svid equals 200.
- Packet encapsulation mode between N-PEs is RAW (Ethernet) mode.
- Connection between U-PE and N-PE1 is PW connection.
- It requires that establish a layer 2 VPN1 by configuring VPLS, making CE1 and CE2 access each other.

10.5.3.2.2 Group Network Diagram

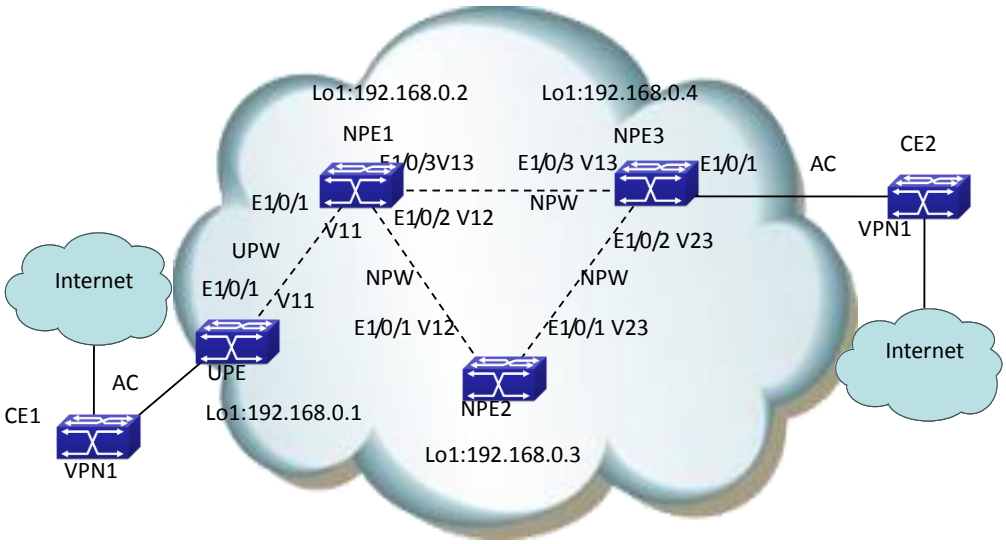


Figure 10-30 H_VPLS network in LSP access mode

10.5.3.2.3 Configuration Steps

Please configure IP address and mask as shown in the above figure for each interface, including VLAN and Loopback interface. Specific configuration steps are omitted. And this is the configuration only for H-VPLS in LDP.

(1) UPE configuration

#Create PW template

```
UPE(config)#pw-class c1
UPE(config-class)#transport-mode ethernet
UPE(config-class)#exit
```

#Configure the routing

```
UPE(config)#router ospf
UPE(config-router)#router-id 192.168.0.1
UPE(config-router)#network 0.0.0.0/0 area 0
UPE(config-router)#exit
```

#Configure MPLS capability and label switching capability (Interface connected with CE does not enable label switching capability)

```
UPE(config)#mpls enable
UPE(config)#int vlan 11
UPE(config-if-vlan11)#label-switching
UPE(config-if-vlan11)#exit
```

#Configure LDP

```
UPE(config)#router ldp
UPE(config-router)#router-id 192.168.0.1
UPE(config-router)#targeted-peer 192.168.0.2
UPE(config-router)#exit
UPE(config)#int vlan 11
UPE(config-if-vlan11)#ldp enable
UPE(config-if-vlan11)#exit
```

Configue instance V1 and remote PE in LDP

```
UPE(config)#vfi v1 100
UPE(config-vfi)#peer 192.168.0.2 pw-class c1
UPE(config-vfi)#exit
```

#Bind VPLS instance mode as Ethernet on port

```
UPE(config-if-ethernet1/0/1)#xconnect vfi 100 mode ethernet
```

(2) NPE1 configuration

#Create PW template

```
NPE1(config)#pw-class c1
```

```
NPE1(config-class)#transport-mode ethernet
NPE1(config-class)#exit

#Configure the routing
NPE1(config)#router ospf
NPE1(config-router)#router-id 192.168.0.2
NPE1(config-router)#network 0.0.0.0/0 area 0
NPE1(config-router)#exit

# Configure MPLS capability and label switching capability
NPE1(config)#mpls enable
NPE1(config)#int vlan 11
NPE1(config-if-vlan11)#label-switching
NPE1(config-if-vlan11)#exit
NPE1(config)#int vlan 12
NPE1(config-if-vlan12)#label-switching
NPE1(config-if-vlan12)#exit
NPE1(config)#int vlan 13
NPE1(config-if-vlan13)#label-switching
NPE1(config-if-vlan13)#exit

#Configure LDP
NPE1(config)#router ldp
NPE1(config-router)#router-id 192.168.0.2
NPE1(config-router)#targeted-peer 192.168.0.1
NPE1(config-router)#targeted-peer 192.168.0.3
NPE1(config-router)#targeted-peer 192.168.0.4
NPE1(config-router)#exit
NPE1(config)#int vlan 11
NPE1(config-if-vlan11)#ldp enable
NPE1(config-if-vlan11)#exit
NPE1(config)#int vlan 12
NPE1(config-if-vlan12)#ldp enable
NPE1(config-if-vlan12)#exit
NPE1(config)#int vlan 13
NPE1(config-if-vlan13)#ldp enable
NPE1(config-if-vlan13)#exit

#Configue instance V1 and remote PE in LDP (peer in UPE needs to disables level division)
NPE1(config)#vfi v1 100
NPE1(config-vfi)#peer 192.168.0.3 pw-class c1
NPE1(config-vfi)#peer 192.168.0.4 pw-class c1
NPE1(config-vfi)#peer 192.168.0.1 no-split-horizon pw-class c1
NPE1(config-vfi)#exit
```

(3) NPE2 configuration
It is similar to NPE1, the difference is no peer, it does not disable level division

(4) NPE3 configuration
#Create PW template
NPE3(config)#pw-class c1
NPE3(config-class)#transport-mode ethernet
NPE3(config-class)#exit

#Configure the routing
NPE3(config)#router ospf
NPE3(config-router)#router-id 192.168.0.4
NPE3(config-router)#network 0.0.0.0/0 area 0
NPE3(config-router)#exit

Configure MPLS capability and label switching capability
NPE3(config)#mpls enable
NPE3(config)#int vlan 13
NPE3(config-if-vlan11)#label-switching
NPE3(config-if-vlan11)#exit
NPE3(config)#int vlan 23
NPE3(config-if-vlan12)#label-switching
NPE3(config-if-vlan12)#exit

#Configure LDP
NPE3(config)#router ldp
NPE3(config-router)#router-id 192.168.0.4
NPE3(config-router)#targeted-peer 192.168.0.2
NPE3(config-router)#targeted-peer 192.168.0.3
NPE3(config-router)#exit
NPE3(config)#int vlan 13
NPE3(config-if-vlan11)#ldp enable
NPE3(config-if-vlan11)#exit
NPE3(config)#int vlan 23
NPE3(config-if-vlan12)#ldp enable
NPE3(config-if-vlan12)#exit

#Configue instance V1 and remote PE in LDP
NPE3(config)#vfi v1 100
NPE3(config-vfi)#peer 192.168.0.2 pw-class c1
NPE3(config-vfi)#peer 192.168.0.3 pw-class c1
NPE3(config-vfi)#exit

#Bind VPLS instance on port
N-PE3(config-if-ethernet1/0/1)#xconnect vfi 100 mode vlan svid 200

10.5.3.3 Access H-VPLS with QinQ

10.5.3.3.1 Network Requirement

- CE1 and CE2 belong to two stations, while they belong to the same VPN1
- CE1 and CE2 can access to PE via port Eth1/0/2 and Eth1/0/1 separately
- U-PE is an access device of user, it accesses N-PE1 with layered QinQ mode, Svid is 100
- CE2 uses the normal access mode as VLAN, Svid is 200.
- Packet encapsulation mode between N-PEs is Tagged mode.
- Connection between U-PE and N-PE1 is QinQ connection.
- It requires that establish a layer 2 VPN1 by configuring VPLS, making CE1 and CE2 access each other.

10.5.3.3.2 Group Network Diagram

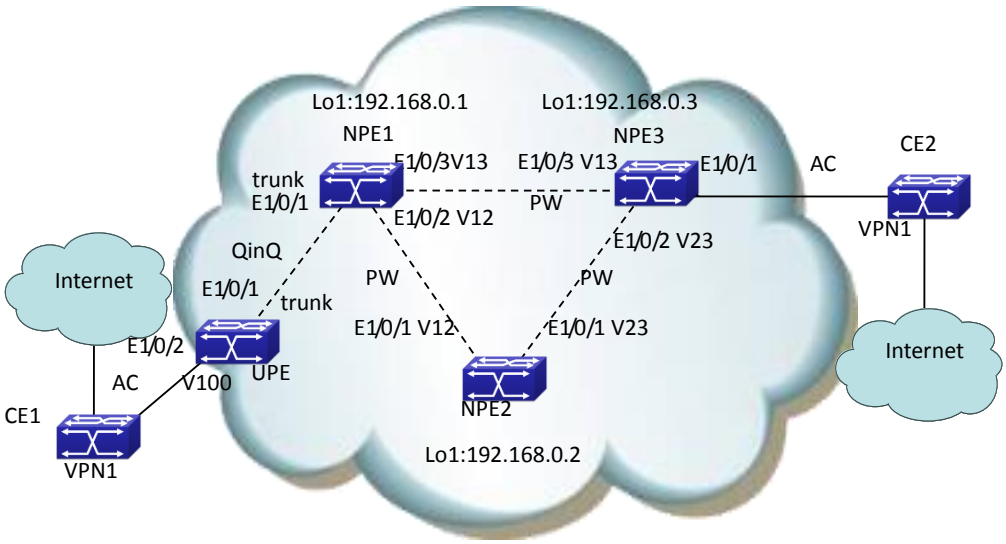


Figure 10-31 H_VPLS network in QinQ access mode

10.5.3.3.3 Configuration Steps

Please configure IP address and mask as shown in the above figure for each interface, including VLAN and Loopback interface. Specific configuration steps are omitted. And this is the configuration only for H-VPLS in QinQ mode.

(1) UPE configuration

#Enable QinQ on port, VLAN of this port belongs to is 100

UPE(config-if-ethernet1/0/2)#switchport access vlan 100

UPE(config-if-ethernet1/0/2)#dot1q-tunnel enable

#Configure the port connected with NPE1 as Trunk, trunk allows all VLANs to pass by default
UPE(config-if-ethernet1/0/1)#switchport mode trunk

(2) NPE1 configuration
#Create PW template
NPE1(config)#pw-class c1
NPE1(config-class)#transport-mode Vlan
NPE1(config-class)#exit

#Configure the routing
NPE1(config)#router ospf
NPE1(config-router)#router-id 192.168.0.1
NPE1(config-router)#network 0.0.0.0/0 area 0
NPE1(config-router)#exit

Configure MPLS capability and label switching capability
NPE1(config)#mpls enable
NPE1(config)#int vlan 12
NPE1(config-if-vlan12)#label-switching
NPE1(config-if-vlan12)#exit
NPE1(config)#int vlan 13
NPE1(config-if-vlan13)#label-switching
NPE1(config-if-vlan13)#exit

#Configure LDP
NPE1(config)#router ldp
NPE1(config-router)#router-id 192.168.0.1
NPE1(config-router)#targeted-peer 192.168.0.2
NPE1(config-router)#targeted-peer 192.168.0.3
NPE1(config-router)#exit
NPE1(config)#int vlan 12
NPE1(config-if-vlan12)#ldp enable
NPE1(config-if-vlan12)#exit
NPE1(config)#int vlan 13
NPE1(config-if-vlan13)#ldp enable
NPE1(config-if-vlan13)#exit

Configue instance V1 and remote PE in LDP
NPE1(config)#vfi v1 100
NPE1(config-vfi)#peer 192.168.0.2 pw-class c1
NPE1(config-vfi)#peer 192.168.0.3 pw-class c1
NPE1(config-vfi)#exit

#Bind VPLS instance mode as VLAN on port with UPE QinQ access mode, Svid is100

NPE1(config-if-ethernet1/0/2)#switchport mode trunk
NPE1 (config-if-ethernet1/0/2)#xconnect vfi 100 mode vlan svid 100

(3) NPE2 configuration
It is similar to NPE1

(4) NPE3 configuration
#Create PW template
NPE3(config)#pw-class c1
NPE3(config-class)#transport-mode Vlan
NPE3(config-class)#exit

#Configure the routing
NPE3(config)#router ospf
NPE3(config-router)#router-id 192.168.0.3
NPE3(config-router)#network 0.0.0.0/0 area 0
NPE3(config-router)#exit

Configure MPLS capability and label switching capability
NPE3(config)#mpls enable
NPE3(config)#int vlan 13
NPE3(config-if-vlan11)#label-switching
NPE3(config-if-vlan11)#exit
NPE3(config)#int vlan 23
NPE3(config-if-vlan12)#label-switching
NPE3(config-if-vlan12)#exit

#Configure LDP
NPE3(config)#router ldp
NPE3(config-router)#router-id 192.168.0.3
NPE3(config-router)#targeted-peer 192.168.0.1
NPE3(config-router)#targeted-peer 192.168.0.2
NPE3(config-router)#exit
NPE3(config)#int vlan 13
NPE3(config-if-vlan11)#ldp enable
NPE3(config-if-vlan11)#exit
NPE3(config)#int vlan 23
NPE3(config-if-vlan12)#ldp enable
NPE3(config-if-vlan12)#exit

Configue instance V1 and remote PE in LDP
NPE3(config)#vfi v1 100
NPE3(config-vfi)#peer 192.168.0.1 pw-class c1
NPE3(config-vfi)#peer 192.168.0.2 pw-class c1

NPE3(config-vfi)#exit

#Bind VPLS instance on port
N-PE3(config-if-ethernet1/0/1)#xconnect vfi 100 mode vlan svid 200

10.5.3.4 VPWS Configuration

10.5.3.4.1 Network Requirement

- CE1 and CE2 belong to two stations, while they belong to the same VPN1
- CE1 and CE2 can access to PE via port Eth1/0/1
- CE1 can access to PE1 in Ethernet mode, while CE2 can access to PE2 in VLAN mode, Svid is 200
- Packet encapsulation mode between PE1 and PE2 is RAW (Ethernet) mode
- It requires to establish a layer 2 VPN1 by configuring VPLS, make CE1 and CE2 can access each other

10.5.3.4.2 Group Network Diagram

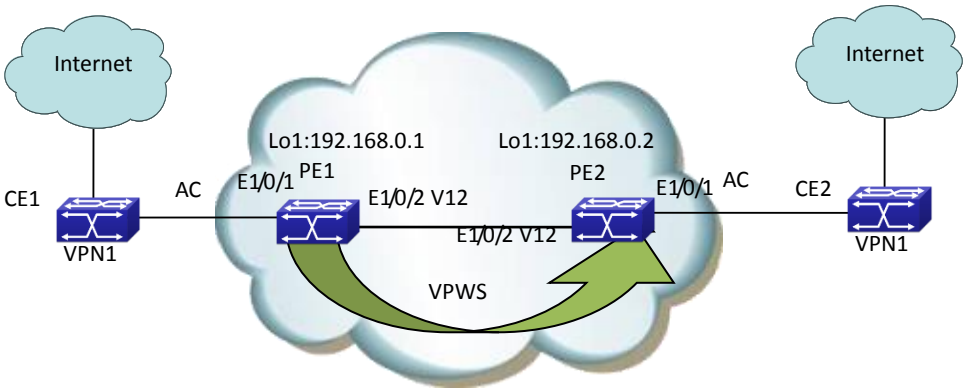


Figure 10-32 VPWS group network diagram

10.5.3.4.3 Configuration Steps

Please configure IP address and mask as shown in the above figure for each interface, including VLAN and Loopback interface. Specific configuration steps are omitted. And this is the configuration only for VPWS.

(1) PE1 configuration

#Create PW template

PE1(config)#pw-class c1

PE1(config-class)#transport-mode ethernet

PE1(config-class)#exit

#Configuration the routing
PE1(config)#router ospf
PE1(config-router)#router-id 192.168.0.1
PE1(config-router)#network 0.0.0.0/0 area 0
PE1(config-router)#exit

#Configure MPLS capability and label switching capability (Interface connected with CE does not enable label switching capability in non-H-VPLS)
PE1(config)#mpls enable
PE1(config)#int vlan 12
PE1(config-if-vlan12)#label-switching
PE1(config-if-vlan12)#exit

#Configure LDP
PE1(config)#router ldp
PE1(config-router)#router-id 192.168.0.1
PE1(config-router)#targeted-peer 192.168.0.2
PE1(config-router)#exit

#Configure remot PE of VPWS
PE1(config)#l2-vc 192.168.0.2 pw-id 1 pw-class c1

#Bind VPWS mode as ethernet on port
PE1(config-if-ethernet1/0/1)#xconnect l2-vc pw-id 1 mode Ethernet

(2) PE2 configuration
#Create PW template
PE2(config)#pw-class c1
PE2(config-class)#transport-mode vlan
PE2(config-class)#exit

#Configure the routing
PE2(config)#router ospf
PE2(config-router)#router-id 192.168.0.2
PE2(config-router)#network 0.0.0.0/0 area 0
PE2(config-router)#exit

#Configure MPLS capability and label switching capability (Interface connected with CE does not enable label switching capability in non-H-VPLS)
PE2(config)#mpls enable
PE2(config)#int vlan 12
PE2(config-if-vlan12)#label-switching
PE2(config-if-vlan12)#exit

#Configure LDP
PE2(config)#router ldp
PE2(config-router)#router-id 192.168.0.2
PE2(config-router)#targeted-peer 192.168.0.1
PE2(config-router)#exit
PE2(config)#int vlan 12
PE2(config-if-vlan12)#ldp enable
PE2(config-if-vlan12)#exit

#Configure remote PE of VPWS
PE2(config)#l2-vc 192.168.0.1 pw-id 1 pw-class c1

#Bind VPWS mode as VLAN, Svid is 200
PE2(config-if-ethernet1/0/1)#xconnect l2-vc pw-id 1 mode vlan svid 200

10.5.4 VPLS Troubleshooting

When configure and use VPLS, L2 VPN may not work correctly because of physical connection and error in configuration. So users should note the following factos:

- First, we should ensure that OSPF neighbor among PE1, P and PE2 works correctly, and routing including loopback interface works correctly, that is, check whether all PE devices exist the routing of the peer PE, P.
- Second, we should ensure that whether PE1, P and PE2 globally enable MPLS and LDP or not, at the same time whether they enable Label-switching and LDP at the active interfaces or not. Also we should check whether LDP session on them is correct or not. Furthermore, we should ensure that LDP's remote neighbor, namely targeted peer address, is correct and neighbor is established correctly. At the same time, we should guarantee address of LDP's remote peer is same as that in VFI or L2VC peer.
- Then, in order to check whether PW in PE is established correctly, we can use command **show vpls peer xxx**. The correct PW state is up, or we should check whether PW encapsulation method (Ethernet or vlan), VFI ID, MTU and transmission mode, etc. in two peers are same or not.
- We should judge whether VFI port is bound to correct VFI or not, and whether access mode and Svid are right or not. Additionally if the process is in VPLS mode, we should check whether pw-id bound by port is correct and pw-id bound by PEs are same.
- Finally, in the case of the process are correct, we can use command **show vfi mac-addresses-table** to check MAC address in the local VPN, and both peers can display MAC address of the peer CE.
- Additionally, in the case that MAC address of the peer CE cannot be displayed in CE and flow does not recovered after we restarted devices, please wait for a moment patiently. Because it needs a time to establish connection among OSPF, LDP and PW.

☞ Notice: please do not enable 802.1x, STP or QinQ, etc. after PE port is bound to VFI. Because this may effect user's network. Furthermore, please do not add this port to layer 3 interface in use of public network, because protocols of routing (for example, OSPF and RIP) and multicast (IGMP and MLD) in layer 3 interface that the port belongs to may result in information leakage of providers to CE.

10.6 MAC-in-MAC

10.6.1 MAC-in-MAC Overview

MAC-in-MAC, also known as PBB (Provider Backbone Bridge), is defined in IEEE 802.1ah. MAC-in-MAC is a Layer-2 VPN (Virtual Private Network) technique. It encapsulates the customer MAC in the service provider MAC, transmits the inner MAC as payload, and thus improves the expandability for Ethernet and secures services.

10.6.1.1 Basic Concept of MAC-in-MAC

1. PBBN

A network using MAC-in-MAC is called a provider backbone bridge network (PBBN) or MAC-in-MAC network. For users, a PBBN is a Layer-2 switching network where Layer-2 connections are between different nodes.

2. PBN

A network connecting the PBBN with the customer network is a provider bridge network (PBN). The customer network can connect to the PBBN directly, or through a PBN.

3. MAC-in-MAC packet

A packet encapsulated by MAC-in-MAC is called a MAC-in-MAC packet.

4. BEB

A backbone edge bridge (BEB) is an edge device in the PBBN, like a PE device in an MPLS network. The BEB encapsulates packets from the customer network by using MAC-in-MAC and forwards them to the PBBN, or de-encapsulates MAC-in-MAC packets from the PBBN and forwards them to the customer network.

5. BCB

A backbone core bridge (BCB) is a core device in the PBBN, like a P device in an MPLS network. It forwards MAC-in-MAC packets according to B-MAC and B-VLAN. A BCB device only forwards packets and learns MAC addresses in the backbone network. It does not learn a large

number of customer MAC addresses. In this way, the network deployment costs are reduced, and the PBBN is given better expandability.

6. B-MAC/B-VLAN

When encapsulating a customer packet, a BEB tags the packet with the service provider MAC address (known as backbone MAC address, B-MAC) and service provider VLAN (known as backbone VLAN, B-VLAN). Note that the B-MAC falls into source B-MAC and destination B-MAC. In the PBBN, a BCB forwards MAC-in-MAC packets according to their B-MAC and B-VLAN.

7. Uplink port/downlink port

The port that connects the BEB to the PBBN is the uplink port, and the port that connects the BEB to the customer network is the downlink port. After the packets from the customer network are encapsulated in MAC-in-MAC packets, they are forwarded out from the corresponding uplink ports on the BEB; after the MAC-in-MAC packets from the PBBN are de-encapsulated, they are forwarded out from the corresponding downlink port on the BEB according to the customer MAC.

8. MAC-in-MAC instance and I-SID

In the PBBN, a MAC-in-MAC instance represents a type of services provided by the service provider, and is uniquely identified by a backbone service instance identifier (I-SID).

10.6.1.2 Basic Network Model of MAC-in-MAC

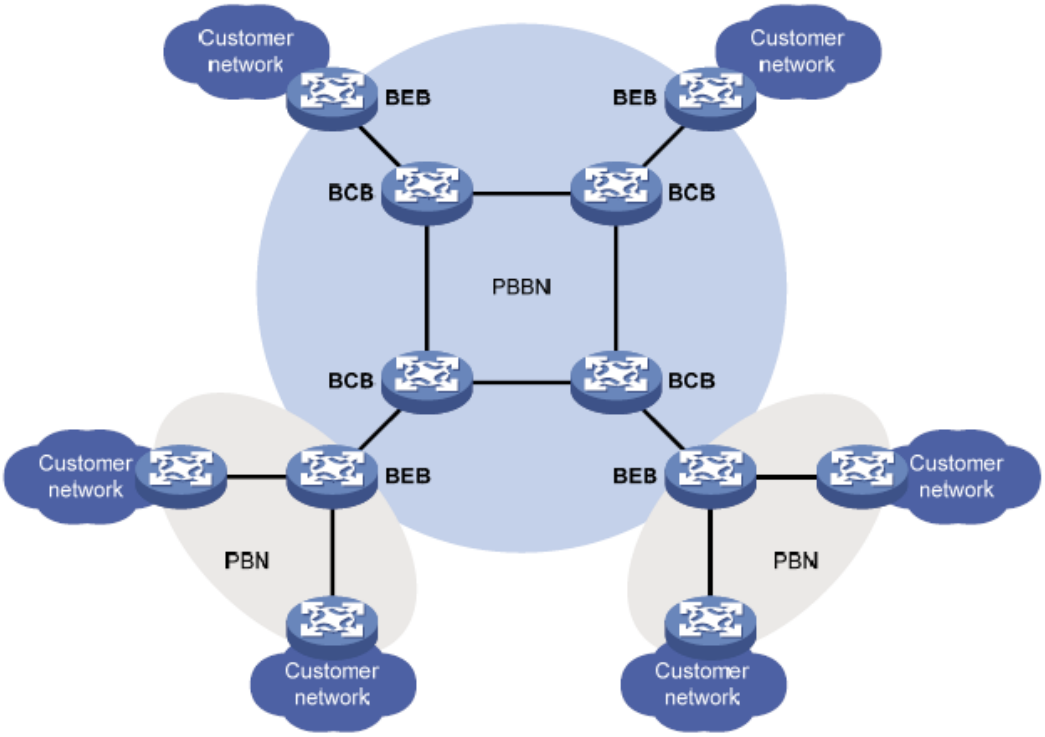


Figure 10-33 Basic Network Model of MAC-in-MAC

MAC-in-MAC model consists of CE, BEB and BCB, the basic concepts are shown in the following.

☞ **CE** (Custom Edge)

Custom Edge device connects with the service provider directly. It may be a router, switch or a host.

☞ **BEB**

A backbone edge bridge (BEB) is an edge device in the PBBN, like a PE device in an MPLS network. The BEB encapsulates packets from the customer network by using MAC-in-MAC and forwards them to the PBBN, or de-encapsulates MAC-in-MAC packets from the PBBN and forwards them to the customer network.

☞ **BCB**

A backbone core bridge (BCB) is a core device in the PBBN, like a P device in an MPLS network. It forwards MAC-in-MAC packets according to B-MAC and B-VLAN. A BCB device only forward packets and learn MAC addresses in the backbone network. It does not learn a large number of customer MAC addresses. In this way, the network deployment costs are reduced, and the PBBN is given better expandability

10.6.1.2.1 Basic Network Model of H-PBBN

H-PBBN（Hierarchical Provider Backbone Bridging Network） enables a PBB network with the lower level to pass through in a PBB network with the higher level, it encapsulate a outer MIM header for MAC-in-MAC packets in a PBB network with the lower level to form double-layer nesting about MIM header

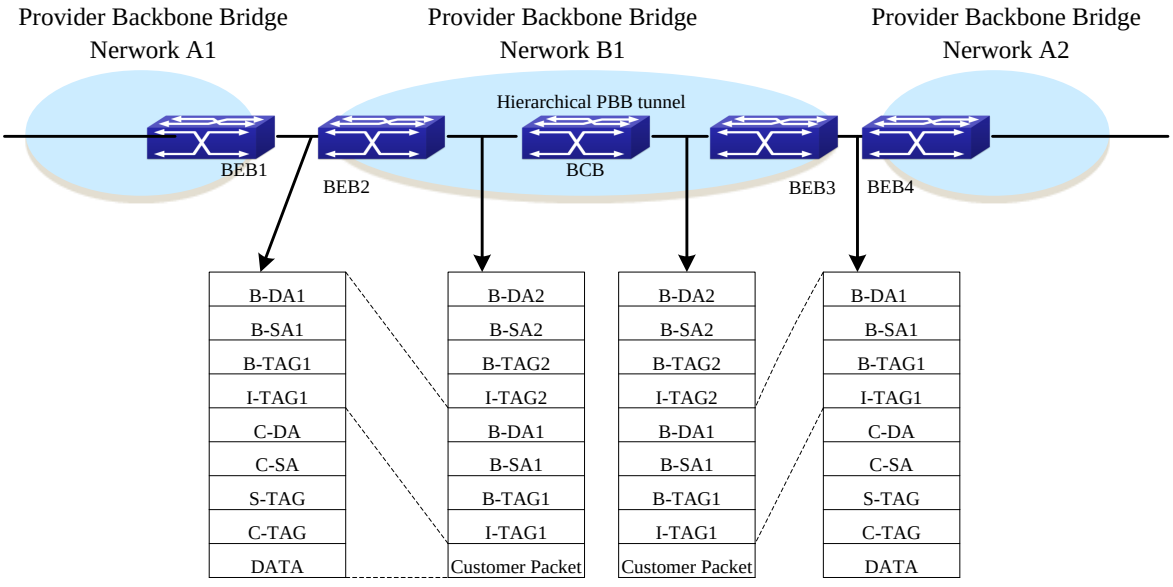


Figure 10-34 Basic Model of H-PBBN

10.6.1.3 Packet Encapsulation of MAC-in-MAC

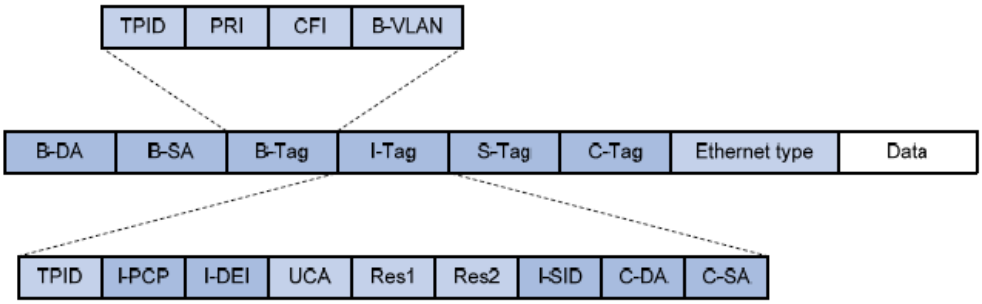


Figure 10-245 Packet encapsulation packet of MAC-in-MAC

Field	Full name	Description
B-DA	Backbone Destination MAC address	Destination B-MAC, outer destination MAC address in a MAC-in-MAC packet. It is the MAC address of the BEB device at the destination end of the PBBN tunnel. The combination of B-DA and B-SA is B-MAC.
B-SA	Backbone Source MAC address	Source B-MAC, outer source MAC address in a MAC-in-MAC packet. It is the MAC address of the BEB device at the source end of the PBBN tunnel. The combination of B-DA and B-SA is B-MAC.
B-Tag	Backbone VLAN Tag	B-VLAN Tag, outer VLAN tag in a MAC-in-MAC packet. It indicates the VLAN information and priority information of the packet within the PBBN. The Tag Protocol Identifier (TPID) in the B-tag is 0x8100.
I-Tag	Backbone Service Instance Tag	Service identifier of a MAC-in-MAC packet. The I-tag contains the backbone service instance priority code point (I-PCP) and backbone service instance drop eligibility indicator (I-DEI) on the BEB, backbone service instance identifier (I-SID), and the C-DA and C-SA of the customer packet. The TPID of the I-tag is 0x88E7.
S-Tag	Service VLAN Tag	Outer VLAN tag of the packet in the PBN, which indicates the VLAN information and priority information of the packet within the PBN.
C-Tag	Customer VLAN Tag	Inner VLAN tag of the packet in the PBN, which indicates the VLAN information and priority information of the packet within the customer network.

Table 0-1 Some key fields of a MAC-in-MAC packet encapsulation format

10.6.1.4 Packet Forwarding of MAC-in-MAC

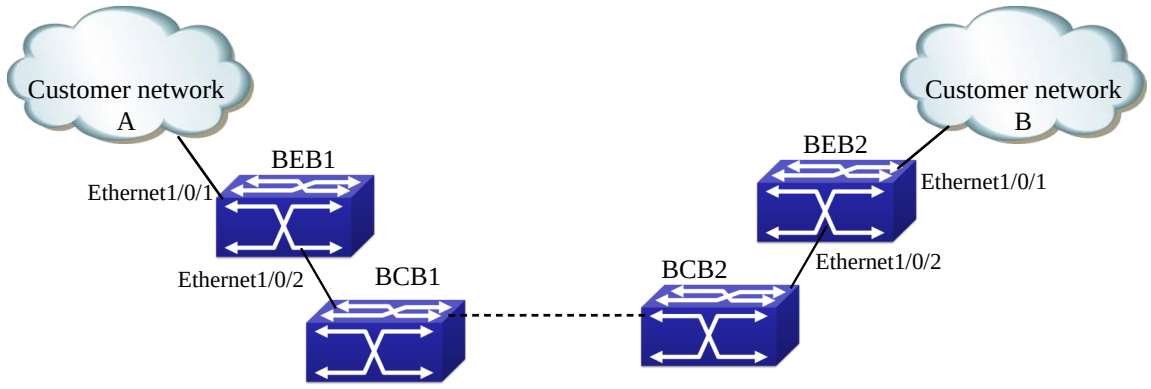


Figure 10-256 PBBN network application

10.6.1.4.1 Access Mode of Downlink Port

There are two access modes on downlink port in the basic model of PBBN, they are VLAN and ethernet, descriptions are shown in the following:

- Access mode of VLAN: Ethernet frame header sent by CE to BEB or BEB to CE with a VLAN TAG, it is a service delimiter added by provider's device to distinguish users, and we call it as S-TAG.
- Access mode of Ethernet: Ethernet frame header sent by CE to BEB or BEB to CE without a service delimiter. If there is a VLAN TAG in it, however, this TAG is an inner one, and it is meaningless to BEB. We call it as C-TAG.

10.6.1.4.2 Packet Forwarding of Ethernet Access

In Fig 0-4, downlink port of BEB uses the access mode of Ethernet, packet forwarding procedures from customer A to customer B are as follows:

1. Customer A sends the packets with U-Tag to BEB1.
2. BEB1 chooses an appropriate B-DA according to the destination MAC address and adds an outer MAC header for the packets.
3. In order to transmit the packets via the public network tunnel, BEB1 adds the public network B-VLAN for the outer MAC header and transmit the packets to BEB2.
4. After BEB2 receives the packets, it finds out the relevant VFI according to B-SA+B-VLAN, and sends the packets with U-TAG to customer B.

10.6.1.4.3 Packet Forwarding of VLAN Access

In Fig 0-4, downlink port of BEB uses the access mode of VLAN, packet forwarding procedures from customer A to customer B are as follows:

1. Packets sent by customer A should be with S-TAG. If packets forwarded by CE1 without VLAN TAG or with unmatched VLAN TAG, it executes the normal layer-2 forwarding process.
2. BEB1 chooses an appropriate B-DA according to the destination MAC address and adds an outer MAC header for the packets.
3. In order to transmit the packets via the public network tunnel, BEB1 adds the public network

B-VLAN for the outer MAC header and transmit the packets to BEB2.
4. After BEB2 receives the packets, it finds out the relevant VFI according to B-SA+B-VLAN, and sends the packets with S-TAG to customer B.

10.6.1.5 MAC-in-MAC Advantages

MAC-in-MAC advantages are shown in the following:

- ☞ MAC-in-MAC solves the limit about QinQ only with 4096 SVLANs.
- ☞ MAC-in-MAC isolates the MAC of the customers, and enhances the security of Ethernet service.
- ☞ MAC-in-MAC extends the space of the MAC addresses.

10.6.2 MAC-in-MAC Configuration

MAC-in-MAC configuration task list:
1. Configure VFI and the relevant parameters (required)
   (1) Create VFI, specify VFI ID and ISID and enter VFI view
   (2) Configure BVLAN for VFI (optional)
   (3) Configure the default multicast group for VFI (optional)
2. Configure downlink port to bind with VFI (required)
   (1) Enter port view
   (2) Bind downlink port to the specified VFI and enter the specified access mode
3. Configure uplink port to bind with VFI (required)
   (1) Enter port view
   (2) Configure a port to be uplink port of VFI

1. Configure VFI and the relevant parameters (required)
   (1) Create VFI, specify VFI ID and ISID and enter VFI view
   (2) Configure BVLAN for VFI (optional)
   (3) Configure the default multicast group for VFI (optional)

Command	Explanation
Global mode	
vfi <*vfi-name*> <*vfi-id*> mim <*i-sid*> **no vfi <*vfi-name*>**	Create a MIM instance and enter VFI mode, at the same time, we must specify the unique MIM instance name, VFI-ID and ISID in global mode. No command deletes the corresponding VFI.
VFI mode	
mim bvlan <*vlan-id*> **no mim bvlan <*vlan-id*>**	Specify a BVLAN for VFI instance, no command restores the default BVLAN.
mim address destination default <*mac-addr*> **no mim address destination default <*mac-add*>**	Specify the remote destination B-DA, no command restores the default value.

2. Configure downlink port to bind with VFI (required)

(1) Enter port view

(2) Bind downlink port to the specified VFI and enter the specified access mode

Command	Explanation
Port mode	
xconnect vfi <*vfi-id*> [mode {ethernet | vlan [svid <*svid*>]}] **no xconnect vfi <*vfi-id*> [mode {ethernet | vlan [svid <*svid*>]}]**	Bind a downlink port with a MIM instance and enter the access mode. No command deletes the MIM instance bound by the port.

3. Configure uplink port to bind with VFI (required)

(1) Enter port view

(2) Configure a port to be uplink port of VFI

Command	Explanation
Port mode	
mim uplink vfi <*vfi-id*> **no mim uplink vfi <*vfi-id*>**	Specify a port to be the uplink port of MAC-in-MAC instance, no command cancels the operation.

10.6.3 Typical Example of MAC-in-MAC

10.6.3.1 Basic Application Scene of MAC-in-MAC

10.6.3.1.1 Network Requirement

- CE1 and CE2 belong to two stations of CustomerA and CustomerB respectively, and they belong to the same VPN1.
- CE1 and CE2 are connected to BEB devices via Eth1/0/1 respectively.
- Access mode of Ethernet is used between CE1 and BEB1, access mode of VLAN is used between CE2 and BEB2, and Svid is set as 200.
- Create layer-2 VPN1 by configuring MAC-in-MAC mode, which enable CE1 and CE2 to access each other.

10.6.3.1.2 Network Diagram

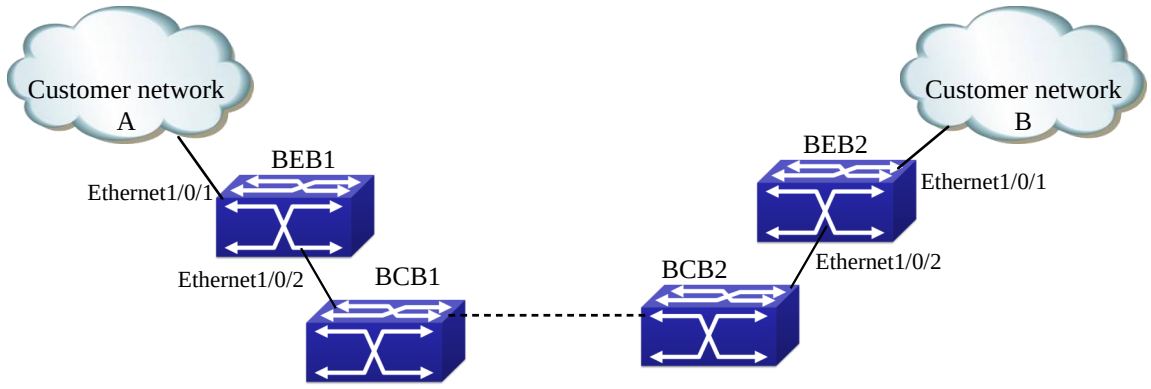


Figure 10-267 PBBN network application

10.6.3.1.3 Configuration Procedures

(1) BEB1 configuration
#Create VFI
BEB1(config)#vfi a 1 mim 17
BEB1(config-vfi)#mim bvlan 2
BEB1(config-vfi)#exit
BEB1(config)#

#Configure the downlink port
BEB1(config)#interface ethernet 1/0/1
BEB1(config-if-ethernet1/0/1)#xconnect vfi 1
BEB1(config-if-ethernet1/0/1)#exit
BEB1(config)#

#Configure the uplink port
BEB1(config)#interface ethernet 1/0/2
BEB1(config-if-ethernet1/0/2)#switchport mode trunk
BEB1(config-if-ethernet1/0/2)#mim uplink vfi 1
BEB1(config-if-ethernet1/0/2)#exit
BEB1(config)#

(2) BEB2 configuration
#Create VFI
BEB1(config)#vfi a 1 mim 17
BEB1(config-vfi)#mim bvlan 2
BEB1(config-vfi)#exit
BEB1(config)#

#Configure the downlink port
BEB1(config)#interface ethernet 1/0/1
BEB1(config-if-ethernet1/0/1)#xconnect vfi 1 mode vlan svid 200

```
BEB1(config-if-ethernet1/0/1)#exit
BEB1(config)#

#Configure the uplink port
BEB1(config)#interface ethernet 1/0/2
BEB1(config-if-ethernet1/0/2)#switchport mode trunk
BEB1(config-if-ethernet1/0/2)#mim uplink vfi 1
BEB1(config-if-ethernet1/0/2)#exit
BEB1(config)#
```

10.6.4 MAC-in-MAC Troubleshooting

When configuring and using MAC-in-MAC, L2 VPN may be abnormity due to the reasons, such as physical connection and false configuration, please pay attention to the following problems.

- First, ensure that physical connection is normal between BEB1, BCB and BEB2.
- Secondly, ensure that ISIDs of the different BEB are same within a VPN when creating VFI, or else communication is failing.
- Then, ensure that uplink port (it must be trunk or hybrid port) and downlink port are added to VFI of BEB.
- Ensure that the configured BVLAN and the native vlan of uplink port must be different in VFI, or else communication is failing.
- If downlink port uses the access mode of VLAN, we should ensure that the tag of the packets received by BEB accords with svid of downlink port, or else communication is failing.
- Ensure that downlink port with tag is added to BVLAN, or else it will result in the packets without tag from downlink port, so as to communication is falling.
- Finally, in the case of the process are correct, we can use command show vfi mac-addresses-table to check MAC address of the local VFI on BEB1 and BEB2, and both peers can show MAC address of the peer CE.

Chapter 11 Virtualization Configuration

11.1 VSF

11.1.1 Overview

11.1.1.1 Introduction of VSF

VSF is using the VSF port to connect several servers and build up a virtual logical facility. The user cans this virtual machine to manage all the physical facilities that link up together.

Traditional district and the data center network is using the multiple layer network topology structure design as shown below in the Fig 1-1. The following are the drawbacks of this type of network:

(1) The network and the server are complicated, thus it will cause the operation efficiency decrease and at higher cost.

(2) No status internet conk out switch will delay the recover time of application and break out time of operation.

(3) Low usage of resources will decrease the return of income (ROI) and increase the capital.

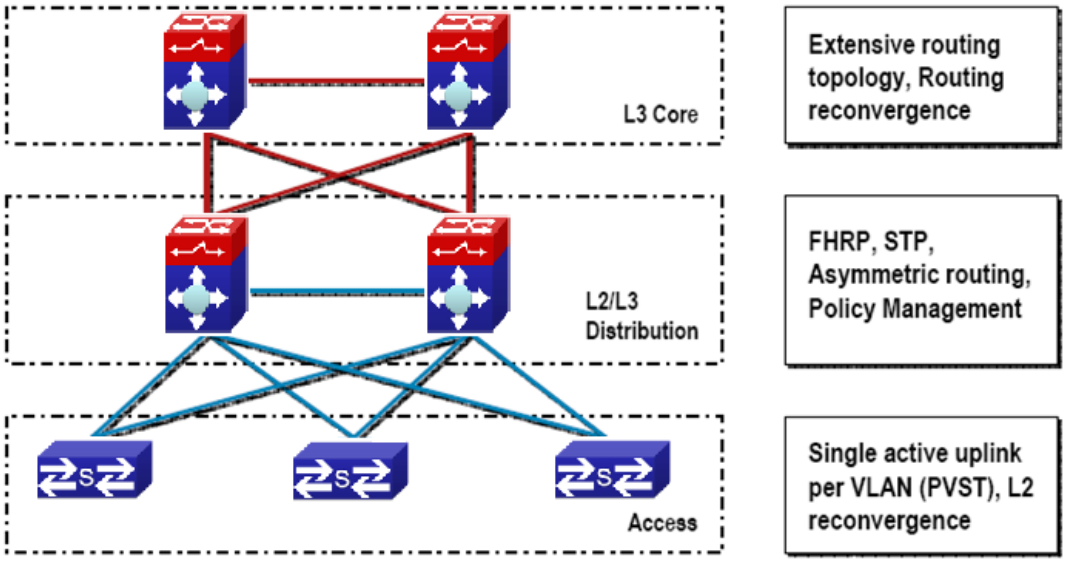


Figure 11-1 Traditional Enterprise Network

In order to solving these problems, VSF comes. Make up more than one device whichsupports VSF to be the single virtualization switch. In VSF, the data panel of the management engine and the exchange array of both switches can be activated at the same time. Members of VSF can throughout the VSL to connection. The VSL standard among the members of the switches is using million-mega Ethernet for connection (maximum have 8 linkages, to provide redundance), likeFig

1-2.

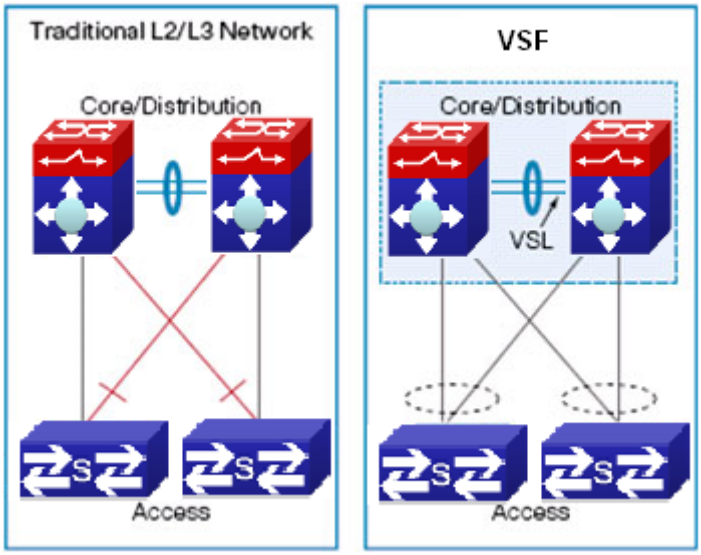


Figure 11-2 VSF

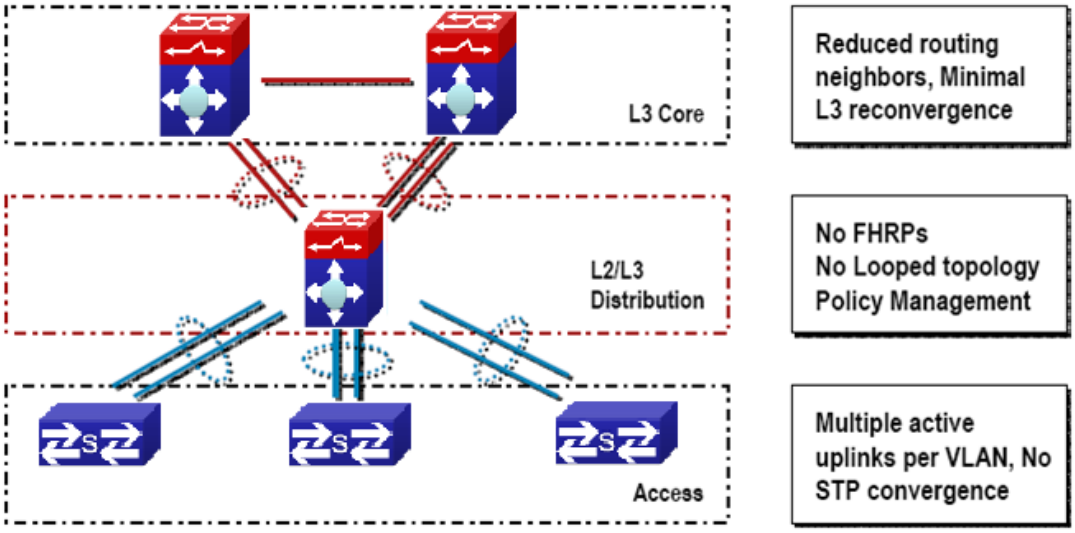


Figure 11-3 Enterprise that adopting virtual technology

Compare with the trandition L2/L3network design, VSF provides with multiple markedness advantages. All in all, we can conclude with 3 main advantages:

1. VSF can increase the efficiency of operation

Only one management point, including configuration documents and gateway IP address (dispense with HSRP/ VRRP/GLBP）

One logical device which is made up by multiple physical devices creates the simple topology without loop and without the spanning tree protocol (STP).

Bottom physical switches connect together by using standard million-mega Ethernet port; it provides a flexible deploy option for the location.

2. VSF can optimize the uninterruptedly communication

Switches status conk out switching will not interrupt the applications which are using the internet status information. With the help of VSF, if one of the virtual switch members is out of order, no need to continous the reconstringency of L2/L3 can restart the virtual switch recover in short time. The active mode must be used for port group.

3. VSF can expand the system bandwidth capacity

In the vsf switch can activate all the L2 bandwidth, can carry on the equilibrium of load on the multiple members of VSF when expanding the bandwidth.

11.1.1.2 Basic Concept

(1) Role

Each of the facilities in the VSF is called member facility. According to the function of each member facility, will divide into 2 different roles:

Master: Responds for manage entire VSF

Standby Master: Is the standby member of VSF, as the Master backup facility for operation. When the Master is out of order, the systems will handover to Standby Master automatically to perform the original job of Master.

Slave: The member facilities that are other than Master and Standby Master in VSF.

Master, Standby Master and Slave are produced by election. For entire VSF, it can only have one Master, one Standby Master, other members should be Slave.

(2) VSF port

It is a logic port which is particular use for VSF. It divides into vsf-port1 and vsf-port2, need to colligate with the VSF physical port and then become efficitive.

(3) VSF physical port

It is the facility that is used for physical connection port for VSF. VSF physical port can be VSF specific port, Ethernet port or light port (different equipment and the model will using different port, please equipt it according to the virtual situtation). Usually, Ethernet port and light port respond for transmitting operation message. After they colligate with the VSF port, they will become the VSF physical port. They will use for transmitting the message within the members. The message including VSF related negotiation message and the operation message that needs to cross over different members.

(4) Merge

Two different VSF operation separately, throughout the physical connection and configuration, it will become one VSF. This process is called merge.



Figure 11-4 Merging of VSF

(5) Split

One VSF is formed. Due to the VSF link is out of order, it will cause the facilities that are border upon cannot connect together physically. Therefore, one stracking will divide into two

stracking. This process is called split.



Figure 11-5 Splitting of tacking

(6) Priority of member

Priority of member is the member facilities attribute, mainly use for the role selection for confirming the role of each members. If the facility has higher the priority, it will have higher chance to become the Master. The default setting of the facilities is 1. If the user want to pre-set particular facility become the Master, he can increase the priority value of that facility manually before build up the VSF.

11.1.1.3 Glossary

VSF is the short form of the Virtual Switching Framework.

VSF Port: VSF port, is a mode on port-channel

VSF AM: the master card of the VSF activities

VSF SM: the backup master card of the VSF activities

VSF Slave: the slave card of VSF, to operate the VSF agreement ,during the switching it means Chassis AM， the not selecte one is VSF AM or VSF SM

VSF Member: the member facilities of VSF

VSF Master Member: the master member facility of VSF

VSF Standby Master Member: the standby master member facility of VSF

VSF Slave Member: the slave member facilities of VSF

HA: High Availability

AM: Active Master, activities controller. It is the switch whose working mode is Master is and it is in the operation status, each of the cassette vsf group can only has a AM at each time.

SM： Standby Master, the working mode in the cassette vsf group is Master and it locates in the standby status.

SSO: Stateful Switchove, the data of in-phase agreement layer will occour the switchover, the backup master agreement apply function which can connect to corresponding function fleetly. The data of this function will not be lose, but the router agreement data is not in-phase

NSF: None Stop Forwarding. Using the independent management characteristics of data layer control, data layer can still transmit data normally as the control layer is out of order

NSF/SSO: Nonstop Forwarding with Stateful Switchover means that undergo the SSO “master-backup interchange”, the data transmit is not affected

Syn process： HA synchronization processes responsable for activities control and backup agreement function collection or distribution and data synchronization communication.

GR: Graceful restart. In order to transmitting uninterruptedly, require the router confer to extend to support the GR ability.

MAD: Multi-Active Detection

11.1.1.4 VSF Typical Application

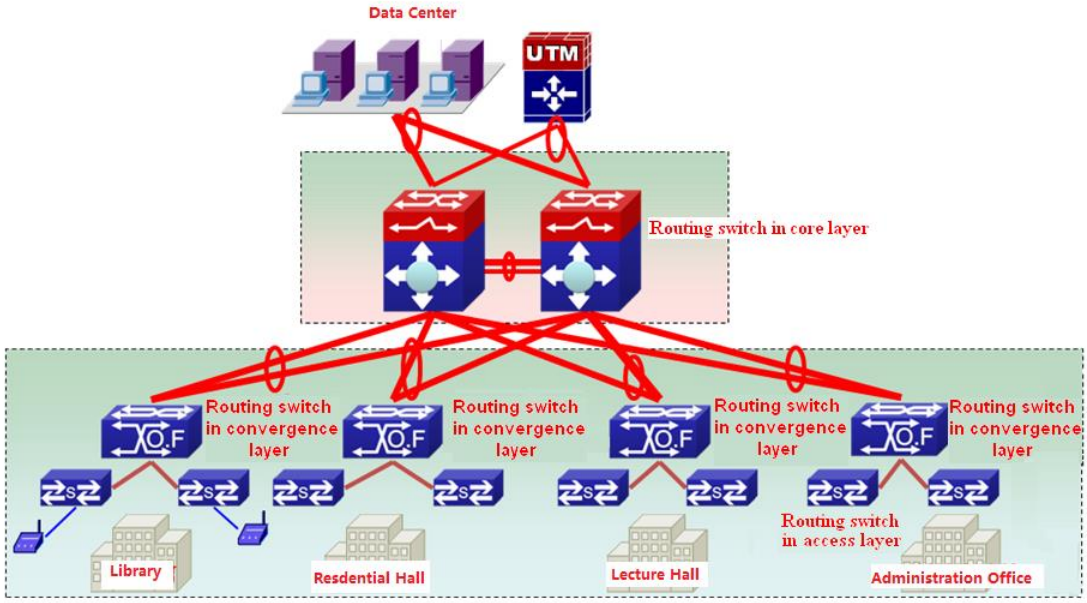


Figure 11-6 application of VSF in the campus data center

Fig 1-6 is the application of VSF in the campus internet. After using the VSF, it groups several facilities together to become the single logic facility and connect to the virtual facility. After the predigestion, the network do not need to use MSTP, VRRP protocol, it simplifies the network configuration. At the same time, it relies on the cross facilities to aggregate the link (The active mode must be used for port group). When some of the members are out of order, it will not rely on MSTP, VRRP protocol etc constringency, thus in increase the reliability.

11.1.1.5 LACP MAD

lacp mad is the dynamic aggregation method based on lacp. Each member device of vsf has one port connecting to the middle device at least.

Notice: The middle device must be the device which supports the extended function of lacp.

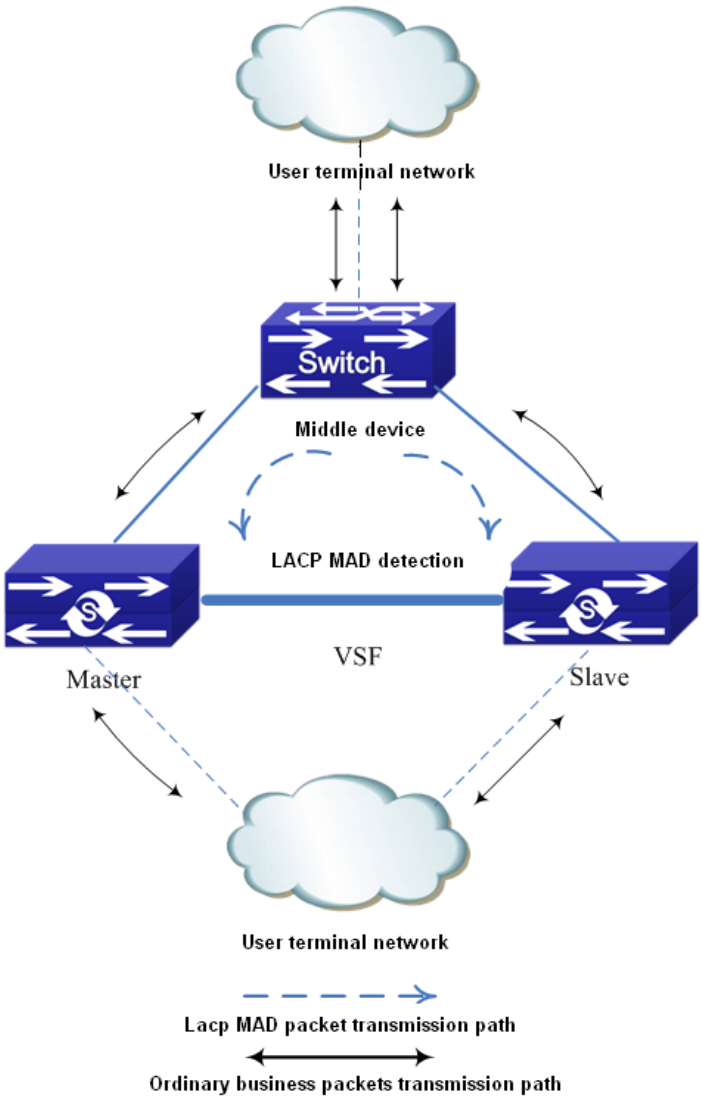


Figure 11-7 LACP MAD detection

LACP MAD detection is achieved through the extended LACP protocol packets content. It defines a new TLV (Type Length Value) in the extended field of LACP protocol packets and this TLV is used for the ActiveID of the interaction VSF. For VSF system, ActiveID value is unique, and it is expressed by the member number of the master device in VSF. After enabling LACP MAD detection, the member device interact the ActiveID information with other member devices through the LACP protocol packets.

When VSF is running normally, the ActiveID values of the LACP protocol packets sent by all member devices are the same, there is no Active conflict.

When the VSF is split into two vsf, the ActiveID values of the LACP protocol packets sent by different VSF member devices are the different and multiple Active conflict will be detected.

11.1.1.6 BFD MAD

The topology of Bfd mad detection is simpler then lacp. But, one vlan is selected for bfd mad detection once, this vlan and the ports in it will be as the specific vlan and ports of bfd mad and

they cannot be configured other functions.

Construction method: select one port on member1 and select one port on member2. Connect them with a line.

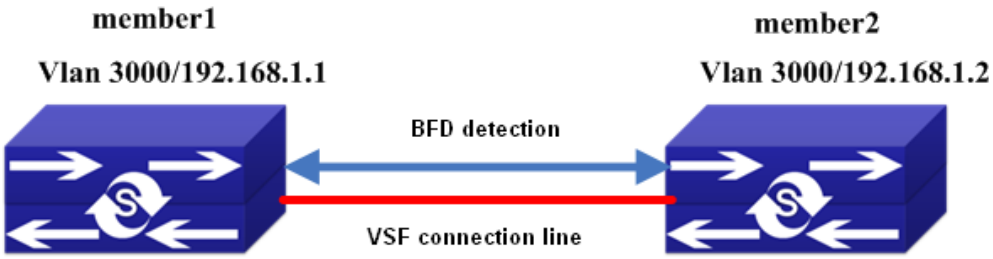


Figure 11-8 BFD MAD detection

BFD MAD detection is achieved through BFD protocol. For BFD MAD detection function running normally, enable BFD MAD detection function under the layer3 interface and configure MAD IP address on this interface. The difference between MAD IP address and the ordinary IP address is that MAD IP address is bond to the member device, each member device in VSF should configured it and they must belong to the same network segment.

When VSF is running normally, only the MAD IP address configured on master is effective, the MAD IP associated with the slave device is not effective, BFD session is under the status of down.

When the VSF is split into two vsf, the mad ip address associated with the member on the new vsf becomes effective. Through the connection line of bfd mad detection, the bfd session will be created between the two vsf to detect the multiple Active conflict.

11.1.2 Relevant VSF Configuration

11.1.2.1 VSF Configuration

VSF configuration task list:

Configure the number of VSF members (necessary)

Configure the priority of VSF members (optional)

Configure the VSF domain (optional)

Configure the logic VSF port

Configure the logic VSF port

Bind the physical port to the logic port

Transform the device from independent operation mode to VSF operation mode

Configure the VSF to merge automatically (optional)

Describe the VSF members (optional)

Configure the down delaying reporting function of the VSF link (optional)

Configure the keeping time of MAC address of VSF group after VSF split (optional)

Transform the device from VSF operation mode to the independent operation mode

Detect the VSF link status' change quickly

Configure the number of VSF members (necessary)

Command	Explanation
Global Mode	
vsf member <*member-id*> **no vsf member <*member-id*>**	Configure/delete the number of VSF members.

Configure the priority and domain of VSF members (optional)

Command	Explanation
Global Mode	
vsf priority <*priority*> **no vsf priority**	Configure/delete the priority of VSF members.
vsf domain <*domain-id*> **no vsf domain**	Configure/delete the VSF domain, the nocommand recovers to be default of 1.

Configure the logic VSF port and Bind the physical port to the logic port

Command	Explanation
Global Mode	
vsf port-group <*port-number*> **no vsf port-group <*port-number*>**	Configure/delete the logic VSF port.
VSF Port Configuration Mode	
vsf port-group interface Ethernet <*interface-list*> **no vsf port-group interface Ethernet <*interface-list*>**	Bind the physical port to the logic VSF port or delete the binding.

Transform the operation modes of the device

Command	Explanation
Global Mode	
switch convert mode (stand-alone | vsf)	Transform the device from independent operation mode to VSF operation mode or transform the device from VSF operation mode to the independent operation mode

Other configurations

Command	Explanation
Global Mode	
vsf auto-merge enable **no vsf auto-merge enable**	Enable the automatic merger function of VSF, the no command disables this function.
vsf member <*member-id*> description <*text*> **no vsf member <*member-id*> description**	Describe the VSF members. This information will be only written in the VSF master configuration file. The no command deletes this information.

vsf link delay<*interval*> **no vsf link delay**	Configure the down delaying reporting function of the VSF link, using for avoid link to split and merge due to changing in short period of time. The **no** command will set the time for delay report to default value.
vsf mac-address persistent <*timer* | *always*> **no vsf mac-address persistent**	Configure VSF split group MAC address retention time. **No** command deletes the MAC address retention time.
vsf non-wait port-inactive **no vsf non-wait port-inactive**	Detect the VSF link status' change quickly for discovering the vsf splitting. The **no** command recovers to be the default method.

11.1.2.2 LACP MAD Configuration

LACP MAD configuration task list:

1. Create the aggregation groups
2. Add the port into the aggregation group, and one is with the active method at least
3. Configure the quick detection (optional)
4. Enable LACP MAD

1. Create the aggregation groups

Command	Explanation
Global Mode	
port-group <*port-group id*> **no Port-group <*port-group id*>**	Create/delete the port-group.

2. Add the port into the aggregation group

Command	Explanation
Global Mode	
interface ethernet <*port-num*>	Enter into the port mode.
Port Mode	
port-group <*port-group id*> mode <*active* | *passive*> **no port-group <*port-group id*>**	Add the port into the aggregation group or let it quit.

3. Configure the quick detection

Command	Explanation
Port Mode	
lacp timeout <*long* | *short*> **no lacp timeout**	Configure/delete the quick detection.

4. Enable LACP MAD

Command	Explanation

Aggregation Port Mode	
**vsf mad lacp <*enable	disable*>**	Enable/disable LACP MAD on port-group.

11.1.2.3 BFD MAD Configuration

BFD MAD configuration task list:

1. Create the vlan used for BFD MAD
2. Add the port used for BFD MAD into the corresponding vlan
3. Configure IP address for BFD MAD layer3 interface
4. Enable BFD MAD function

1. Create the vlan used for BFD MAD

Command	Explanation
Global Mode	
vlan <*vlan-id*> **no vlan <*vlan-id*>**	Configure/delete vlan.

2. Add the port used for BFD MAD into the corresponding vlan

Command	Explanation
VLAN Configuration Mode	
switchport interface ethernet <*port-num*> **no switchport interface ethernet <*port-num*>**	Add the port into vlan or let it out.

3. Configure IP address for BFD MAD layer3 interface

Command	Explanation
Global Mode	
Interface vlan <*vlan-id*>	Enter into the vlan interface configuration mode.
Interface Configuration Mode	
vsf mad ip address <*ip-addr*> <*ip-mask*> member <*member-id*> **no vsf mad ip address <*ip-addr*> <*ip-mask*> member <*member-id*>**	Configure/delete the IP address used for BFD MAD on the layer3 interface.

4. Enable BFD MAD function

Command	Explanation
Interface Configuration Mode	
**vsf mad bfd <*enable	disable*>**	Enable/disable BFD MAD.

11.1.3 Typical VSF Example

Case 1：

Configure under the independent operation mode, let two switches create VSF. The member number of these two facilities is 1 and 2. In order to let vsf member2 as the vsf master, configure the priority of member 2 as 32, and build up two vsf port-group between both facilities. Each of the vsf port-group will colligate one of the million mega port.

switch1 VSF configuration as following：

```
switch1#config
switch1(config)#vsf member 1
switch1(config)#vsf port-group 1
switch1(config-vsf-port1)#vsf port-group interface ethernet 1/1
switch1(config)#vsf port-group 2
switch1(config-vsf-port1)#vsf port-group interface ethernet 1/2
switch1(config)#exit
switch1(config)#switch convert mode vsf
```

switch2 VSF configuration as following：

```
switch2#config
switch1(config)#vsf member 2
switch1(config)#vsf priority 32
switch1(config)#vsf port-group 1
switch1(config-vsf-port1)#vsf port-group interface ethernet 2/1
switch1(config)#vsf port-group 2
switch1(config-vsf-port1)#vsf port-group interface ethernet 2/2
switch1(config)#exit
switch1(config)#switch convert mode vsf
```

Case 2：

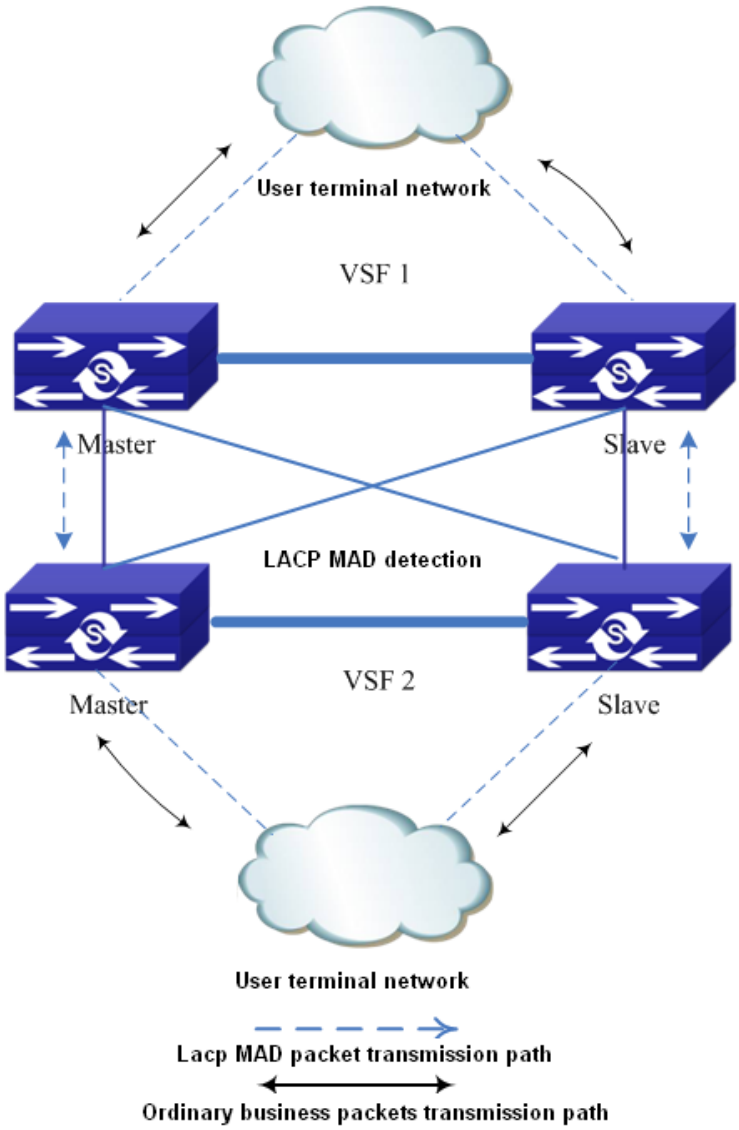


Figure 11-9 lacp mad detection topology

As shown in the above picture, use the lacp mad detection function between two vsf. Vsf1 and vsf2 are the devices which are detected and they are also the middle devices. The configuration is same with above. Proposal: uses create the overlapping connection among the devices to avoid that the vsf1 cannot be the middle device to detect vsf2 after it is split.

Assume that vsf1 and vsf2 use the lacp ports of ethernet1/1/1, ethernet 1/1/2, ethernet 2/1/1 and ethernet 2/1/2.

vsf1 configuration:

Switch(config)#vsf domain 1

Configure vsf domain number, it can be configured as other value but it cannot be the same as vsf2.

Switch(config)#port-group 1

Switch(config)#interface ethernet 1/1/1

Switch(config-if-ethernet1/1/1)#port-group 1 mode active

Switch(config)#interface ethernet 1/1/2

Switch(config-if-ethernet1/1/2)#port-group 1 mode active

Switch(config)#interface ethernet 2/1/1

Switch(config-if-ethernet2/1/1)#port-group 1 mode active

Switch(config)#interface ethernet 2/1/2

Switch(config-if-ethernet2/1/2)#port-group 1 mode active

Switch(config-if-ethernet2/1/2)#interface port-channel 1

Switch(config-if-port-channel1)#vsf mad lacp enable

vsf2 configuration:

Switch(config)#vsf domain 2

Configure vsf domain number, it can be configured as other value but it cannot be the same as vsf1.

Other configuration is same as the configuration on vsf1.

Case 3：

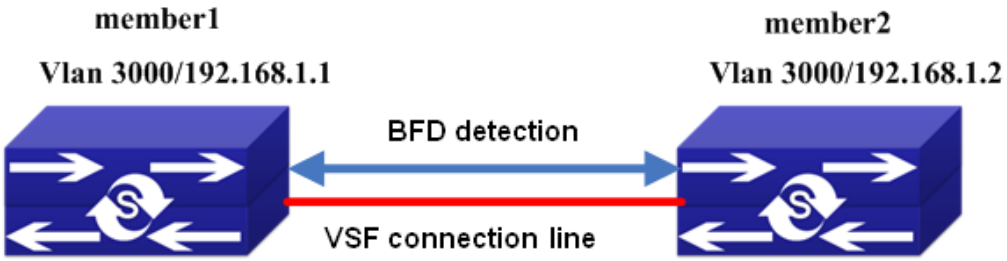


Figure 11-10

As show in the above picture, assume the ports used for bfd mad detection are ethernet1/1/1 and ethernet2/1/1, and the vlan is 3000. The configuration of mad ip address is the network segment of 192.168.1.1 for example.

BFD MAD configuration of switch:

Switch(config)#vlan 3000

Switch(config-vlan3000)#interface ethernet 1/1/1

Switch(config-if-ethernet1/1/1)#switchport access vlan 3000

Switch(config-if-ethernet1/1/1)#interface ethernet 2/1/1

Switch(config-if-ethernet2/1/1)#switchport access vlan 3000

Switch(config-if-ethernet2/1/1)#interface vlan 3000

Switch(config-if-vlan3000)#vsf mad bfd enable

Switch(config-if-vlan3000)#vsf mad ip add 192.168.1.1 255.255.255.0 member 2

Switch(config-if-vlan3000)#vsf mad ip add 192.168.1.2 255.255.255.0 member 1

Notice: In this time, the whole vsf is a device, so BFD MAD configuration is equivalent to conduct on one device.

11.1.4 VSF Troubleshooting

For VSF, under the configuration and usage, the command may not be workable, please pay attention to the following items:

- ☞ Whether it is at the operation mode, because some of the commands can only be configured at VSF operation mode, but some of them can operation at both VSF and independent operation mode.

Under the failure of VSF during the VSF processes, or appear any abnormity, please pay attention to the following thing:

- ☞ First, need to check whether physical connection is correct. Right now, the group VSF machine can only support million-mega port or logic port colligate. Please check whether the colligate port is the million-mega port.
- ☞ Whether the Vsf member id is clash, if the member ids are clashing, the facilities cannot form vsf.
- ☞ Whether the Vsf domain numbers are same. Only the vsf domain numbers are identical, the facilities can form the vsf.
- ☞ To see whether the logical port has any physical connection, for example, double speed, bandwidth restriction, safety authentication, ACL etc configuration.

Under the VSF operation mode to undergo configuration, please pay attention to the following things:

- ☞ During the VSF operation mode, VSF related configuration can be done on the each member separately. However, the configuration done on the members cannot be saved independently. It cans still excute the write function on the vsf master. All the membes' personal configuration that is related to vsf will be save in the main computer's vsf.cfg.
- ☞ Under the VSF operation mode, some of the commands such as VSF domain number, priority of member, member id etc can still be configured or amended. After the configuration, the show run displays as the latest value, it can only be effective as you restart the computer after saving.
- ☞ Under the VSF operation mode, the active mode must be used for port group for ensuring the fault convergence time.

For LACP MAD, under the configuration and usage, if there are problems, please pay attention to the following items:

- ☞ When creating the aggregation group, the aggregation of the device which is detected and the middle device must be identical.
- ☞ The ports of the device which is detected and the middle device cannot be added into the aggregation with on mode for LACP MAD, one should be active method at least.
- ☞ When the device which is detected and the middle device are both vsf, the domain id (vsf domain number) of the two vsf cannot be the same.

For BFD MAD, under the configuration and usage, if there are problems, please pay attention to the following items:

- ☞ bfd mad function cannot be enabled under interface vlan 1.
- ☞ bfd mad function and the anti-ring function are mutually exclusive (such as stp, mstp, etc). if this vsf is configured anti-ring function, the ports used for bfd mad detection should be disabled the anti-ring function, otherwise, the detection may fail.
- ☞ If there is port configured as truck port on this vsf (include port-channel port), please ensure whether the vlan used for bfd mad detection is in this trunk (trunk port belongs to all vlan as default). If it is in, the vlan used for bfd mad detection must be filtered under this port, otherwise, the loop may appears.
- ☞ If there is port configured as hybrid port on this vsf, please ensure that this hybrid port does not belong the vlan used for bfd mad detection, otherwise, the loop may appears.
- ☞ The vlan interface and all the ports in this vlan used for bfd mad detection cannot be configured other configuration except bfd mad detection.
- ☞ If the bfd mad function will not be used any more, clear the configuration and the connection line used for bfd mad should be removed.
- ☞ If bfd mad configuration has been already completed, and other ports, vlan or the command under the global mode should be modified, please ensure whether it conforms the above items.
- ☞ Suggestion: user should select a vlan which is not used in business on the whole as the specific vlan for bfd mad detection.

Chapter 12 openflow Configuration

12.1 Openflow

12.1.1 Introduction to Openflow

1.1 First published in April 2008, Openflow is presented on paper OpenFlow: enabling innovation in campus networks, which is Professor Nick McKeown's achievements exhibition on ACM Communications Review, and it discussed the details of principles of OpenFlow. From the thesis topic, the initial point of departure of OpenFlow is designed for the campus network researchers to test their innovative network architecture, protocols. Taking into account the actual network verify the actual network of innovative ideas better, and researchers cannot modify the access network equipment, the paper proposed separation on OpenFlow control forwarding architecture. It leads out the control logic from the network device, so that researchers can program the control logic to achieve any new network protocols, topologies without changing the network equipment itself. The idea has been first applied in the US GENI research project, implements an innovation experiment platform end to end from a host to a network.

The idea of Openflow is very simple. Network equipment (openflow switch) maintains one or several FlowTable and only forwards according to FlowTable. The generation, maintenance, forwarding of FlowTable are fully achieved by the external controller. But each field can be common configured. The Network operators decide to use of what kind of granularity stream. For example, Network operators only need to rout according to the destination IP. So only the destination IP field is effective in the flow table. The others are all common configuration.

For L2 switching devices, this separation of control and forwarding architecture means that MAC address learning is achieved by the controller. VLAN and basic L3 routing configuration are also issued by controller to switch. For L3 equipment, various types of IGP/EGP routing are running upon the controller. The controller issues to the corresponding router as needed. Sending the flow table is either active or passive. Under the active mode, the controller will issues the flow table information that it gathered to network equipment. And then the network equipment can forward based directly on flow table. Passive mode refers to that when network device receives a packet and it does not match any FlowTable record, the packet will be forwarded to the controller. The controller will decide how to forward the packet and issue the corresponding flow table. The benefit of passive mode is that the network equipment does not need to maintain the entire flow table, the network device gets from the controller for flow table record and stores only when the actual flow is generated. When the aging timer is timeout, the corresponding flow table will be deleted. So the TCAM space is greatly reduced.

From version 0.2.0 to version 1.3.2, Openflow itself has gone through several versions of evolution. Currently version 1.0 and 1.3 are widely used as the stable version for development. In

fact, ONF slowed the progress of evolution of the standard recently. Each version has a different extant of change, our switch supports the openflow 1.0 now.

12.1.2 Openflow Configuration

Openflow configuration task list is as below:

1. Configure to create the TCP connection between openflow switch and controller
2. Configure the openflow vlan or openflow port
3. Associate the openflow card with the openflow vlan or the vlan that the openflow port is in
4. Configure the upper limit of the packets that the switch sends to the controller per second
5. Enable debug

1. Configure to create the TCP connection between openflow switch and controller

Command	Explanation
Global Mode	
openflow mode	Enter into the openflow configuration mode.
Openflow Mode	
openflow tcp A.B.C.D <1-65535> [auxiliary] **no openflow tcp A.B.C.D <1-65535> [auxiliary]**	Create the TCP connection between the openflow switch and the openflow controller appointed by the command (including the IP address and TCP monitoring ports information of the controller). The no command is to disconnect it.

2. Configure the openflow vlan or openflow port

Command	Explanation
Global Mode	
openflow vlan <1-4094> **no openflow vlan**	Configure the vlan controlled by the openflow. The no command configures not to control any vlan anymore.
Port Mode	
openflow enable **no openflow enable**	Configure the port as the openflow port, the no command cancels it.

3. Configure the upper limit of the packets that the switch sends to the controller per second

Command	Explanation
Global Mode	
openflow packet rate limit <1-800>	Configure the upper limit of the packets that the openflow switch sends to the controller every second.

4. Enable debug

Command	Explanation
Admin Mode	
debug openflow {event | packet} **no debug openflow {event | packet}**	Enable the debug information of openflow. The no command disables it.

12.1.3 Openflow Examples

Openflow topology:

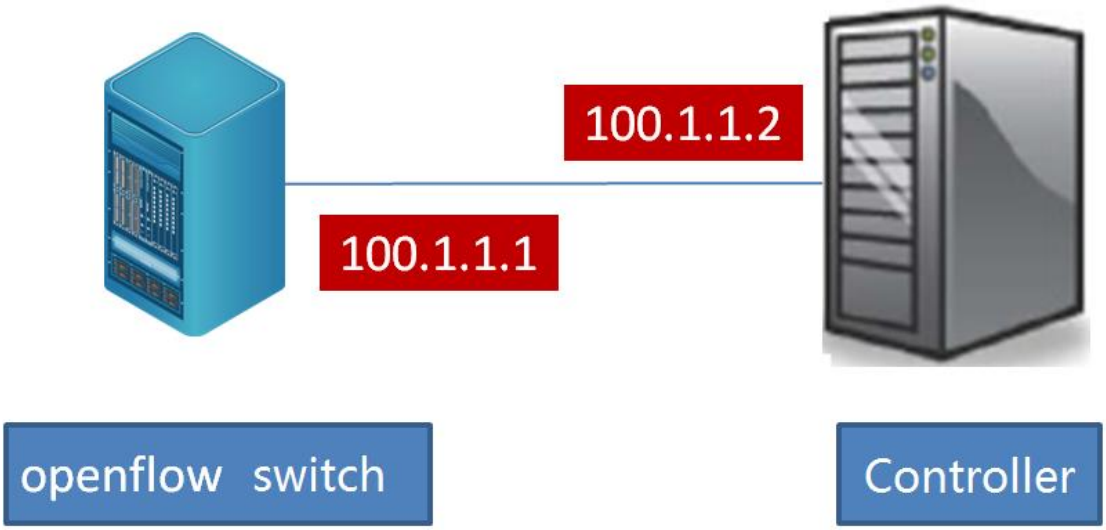


Figure 12-1 Openflow Configuration

In the above figure, the openflow switch and the controller are connected, their ip addresses are shown in the figure.

12.1.3.1 Network Connection between Openflow Switch and Controller

The switch and the controller can be connected through the business interface or Ethernet0, the two kinds of configuration are as below:

1. Business interface connects switch to controller:

The switch and the controller are connected through the business interface 1/0/8.

Configuration modification:

Step 1: Configure the port 1/0/8 to join vlan100.

Step 2: Configure the interface vlan 100, the layer3 interface address is 100.1.1.1.

Configuration steps:

switch(config)#interface ethernet1/0/8

switch(config-if-ethernet1/0/8)#switchport access vlan 100
switch(config-if-ethernet1/0/8)#exit
switch(config)#interface vlan 100
switch(config-if-vlan100)#ip address 100.1.1.1 255.255.255.0

2. Ethernet0 interface connects switch to controller:

Configuration modification:
Step 1: Configure the address of Ethernet0 as 100.1.1.1.

Configuration steps:
switch(config)#interface ethernet0
switch(config-if-ethernet0)#ip address 100.1.1.1 255.255.255.0

12.1.3.2 Vlan Mode Example

Configuration modification:
Step 1: Configure to create the TCP connection between openflow switch and controller.
Step 2: Configure the openflow vlan.
vlan 10 is configured as openflow vlan.
Step 3: Associate the openflow card with openflow vlan.

Configuration steps:
Switch:
switch(config)# openflow mode
switch(config-openflow)#openflow tcp 100.1.1.2 6633
switch(config-openflow)#exit
switch(config)#openflow vlan 10
switch(config)#openflow-connnect slot 2 vlan 10

12.1.4 Openflow Troubleshooting

→ The vlan mode only supports to configure one vlan as the openflow vlan, the new configured openflow vlan will cover the old configured openflow vlan.
→ The switch openflow module can be viewed the event information through the command of debug openflow events.
→ After the switch openflow module received the packet, it will be conducted, the detailed packet information can be viewed through the command of debug openflow packet.
→ Use the command of show openflow acl all (openflow1.0) to check if the flow rule is issued successfully by switch.
→ Use the command of show tcp to check if the connection is created between switch and controller.

12.2 POX Controller

12.2.1 OPENFLOW Configuration (match)

12.2.1.1 ./pox.py openflow.of_01 --address=x.x.x.x –port=yy py

Command：./pox.py openflow.of_01 --address=x.x.x.x –port=yy py
Function： The controller enables the corresponding address monitoring.
Parameters：address，port，py
Default： None.
Command Mode： The commands including the path of pox.py.
Usage Guide：./pox.py openflow.of_01 --address=x.x.x.x –port=yy py。
Example：
Controller terminal:
Enable the pox address monitoring. The port of 6633 will be used to enter into the interactive mode as default.
root@long-Aspire-4733Z:/home/long/usr/pox# ./pox.py openflow.of_01 --address=6.6.6.6 py
POX 0.1.0 (betta) / Copyright 2011-2013 James McCauley, et al.
INFO:core:POX 0.1.0 (betta) is up.
This program comes with ABSOLUTELY NO WARRANTY. This program is free software,
and you are welcome to redistribute it under certain conditions.
Type 'help(pox.license)' for details.
Ready.
POX>
Pc terminal:
DCRS-6500-52X(config)#openflow mode
DCRS-6500-52X(config-openflow)#openflow tcp 6.6.6.6 6633
Controller terminal:
The switch connection prompt
POX> INFO:openflow.of_01:[00-03-0f-27-5e-91 1] connected

12.2.1.2 Import pox.openflow.libopenflow_01 as of

Command：import pox.openflow.libopenflow_01 as of
Function： Export the core module and name it as of.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：import pox.openflow.libopenflow_01 as of. Export the core module and name it as of.
Example：
Controller terminal:

POX> **import pox.openflow.libopenflow_01 as of**

12.2.1.3 core.openflow.connections.keys()

Command：core.openflow.connections.keys()
Function： Get the key of the “of switch” connected to the controller.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：core.openflow.connections.keys(). Send the msg by the key after got it.
Example：
Controller terminal:
POX> **core.openflow.connections.keys()**

12.2.1.4 core.openflow.connections[key].send(msg)

Command：core.openflow.connections[key].send(msg)
Function： Send the msg through the specific key.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：core.openflow.connections[key].send(msg). The **key** is the switch code which was got by the last command.
Example：
Controller terminal:
POX> **core.openflow.connections[key].send(msg)**

12.2.1.5 msg=of.ofp_flow_mod()

Command：msg=of.ofp_flow_mod()
Function： Compile the message and the type of message is flow mod.
Parameters：command： 0 means to ADD (add flow). 1 means to MODIFY, 2 means to MODIFY_STRICT, 3 means to DELETE (delete all the flow rules) and 4 means to DELETE_STRICT (delete the flow rules according to the mask and priority)
Default： command：0
Command Mode： Interactive mode, pox>
Usage Guide：msg=of.ofp_flow_mod(). The type of the msg message is flow mod.
Example：
Controller terminal:
POX> **msg=of.ofp_flow_mod()**

12.2.1.6 msg.priority=x

Command：msg.priority=x

Function：Configure the priority of the rules.
Parameters：None.
Default：None.
Command Mode：Interactive mode, pox>
Usage Guide：**msg.priority=x.** The priority must be initialized and the rage is from 1 to5.
Example：
Controller terminal:
POX> **msg.priority=5**

12.2.1.7 msg.match.in_port=y

Command：**msg.match.in_port=y**
Function：Match the access port.
Parameters：None.
Default：None.
Command Mode：Interactive mode, pox>
Usage Guide：msg.match.in_port=y. y is the value of the field.
Example：
Controller terminal:
Appoint the rule to match the access port of 1.
POX> **msg.match.in_port=1**

12.2.1.8 msg.match.dl_src=EthAddr(“”)

Command：**msg.match.dl_src=EthAddr(“”)**
Function：Match the source mac.
Parameters：None.
Default：None.
Command Mode：Interactive mode, pox>
Usage Guide：msg.match.dl_src=EthAddr(“”)
Example：
Controller terminal:
Appoint the rule to match the source mac of 00:00:00:00:00:11.
POX> **msg.match.dl_src=EthAddr(“00:00:00:00:00:11”)**

12.2.1.9 msg.match.dl_dst=EthAddr(“”)

Command：**msg.match.dl_dst=EthAddr(“”)**
Function：Match the destination mac.
Parameters：None.
Default：None.
Command Mode：Interactive mode, pox>
Usage Guide：msg.match.dl_dst=EthAddr(“”)
Example：

Controller terminal:
Appoint the rule to match the destination mac of 00:00:00:00:00:11.
POX> **msg.match.dl_dst=EthAddr("00:00:00:00:00:11")**

12.2.1.10 msg.match.dl_type=x

Command：msg.match.dl_type=x
Function： Match the type of ethernet.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide： msg.match.dl_type=
Example：
Controller terminal:
Appoint the rule to match the packets of the ip type.
POX> **msg.match.dl_type=0x800**

12.2.1.11 msg.match.dl_vlan=x

Command：msg.match.dl_vlan=x
Function： Match the vlan id.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide： msg.match.dl_vlan=
Example：
Controller terminal:
Appoint the rule to match the vlan id.
POX> **msg.match.dl_vlan=3**
Explanation: dl_vlan must be "of vlan".

12.2.1.12 msg.match.dl_vlan_pcp=x

Command：msg.match.dl_vlan_pcp=x
Function： Match the tos value.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide： msg.match.dl_vlan_pcp=
Example：
Controller terminal:
Appoint the rule to match the cos value.
POX> **msg.match.dl_vlan_pcp=3**
Explanation: dl_vlan_pcp must be from 0 to 7.

12.2.1.13 msg.match.nw_src=

Command：msg.match.nw_src=
Function： Match the source ip address.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide： msg.match.nw_src=
Example：
Controller terminal:
Appoint the rule to match the source IP.
POX> **msg.match.dl_type=0x800**
POX> **msg.match.nw_src="192.168.2.133/24"**
Explanation: The type of ethernet must be appointed and the mask of ip can be appointed with "/".

12.2.1.14 msg.match.nw_dst=

Command：msg.match.nw_dst=
Function： Match the destination ip address.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide： msg.match.nw_dst=
Example：
Controller terminal:
Appoint the rule to match the destination IP address.
POX> **msg.match.dl_type=0x800**
POX> **msg.match.nw_dst="192.168.2.133/24"**
Explanation: The type of ethernet must be appointed and the mask of ip can be appointed with "/".

12.2.1.15 msg.match.nw_proto=x

Command：msg.match.nw_proto=x
Function： Match the protocol type.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide： msg.match.nw_proto=
Example：
Controller terminal:
Appoint the rule to match the packet of IP type.
POX> **msg.match.dl_type=0x800**

POX> **msg.match.nw_proto=6**

Explanation: The type of ethernet must be appointed and then match the ip protocol.

12.2.1.16 msg.match.nw_tos=x

Command：msg.match.nw_tos=x

Function： Match the tos.

Parameters： None.

Default： None.

Command Mode： Interactive mode, pox>

Usage Guide： msg.match.nw_tos=

Example：

Controller terminal:

Appoint the rule to match the ip protocol.

POX> **msg.match.dl_type=0x800**

POX> **msg.match.nw_tos=64**

Explanation: The type of ethernet must be appointed and then match the tos value.

12.2.1.17 msg.match.tp_src=x

Command：msg.match.tp_src=

Function： Match the tcp source port.

Parameters： None.

Default： None.

Command Mode： Interactive mode, pox>

Usage Guide： msg.match.tp_src=

Example：

Controller terminal:

Appoint the rule to match the tcp source port.

POX> **msg.match.dl_type=0x800**

POX> **msg.match.nw_proto=6**

POX> **msg.match.tp_src=179**

Explanation: The type of ethernet must be appointed, then match the ip protocol and match the tcp port at last.

12.2.1.18 msg.match.tp_dst=x

Command：msg.match.tp_dst=

Function： Match the tcp destination port.

Parameters： None.

Default： None.

Command Mode： Interactive mode, pox>

Usage Guide： msg.match.tp_dst=

Example：

Controller terminal:
Appoint the rule to match the tcp destination port.
POX> **msg.match.dl_type=0x800**
POX> **msg.match.nw_proto=6**
POX> **msg.match.tp_dst=179**
Explanation: The type of ethernet must be appointed, then match the ip protocol and match the tcp port at last.

12.2.1.19 msg.idle_timeout=x

Command：msg.idle_timeout=
Function： In the interval of idle, if there is no packet triggering this action, this rule will be deleted.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide： msg.idle_timeout=30
Example：
Controller terminal:
Appoint the idle time of the rule as 30s.
POX> **msg.idle_timeout=30**
Explanation: None.

12.2.1.20 msg.hard_timeout=x

Command：msg.hard_timeout=
Function： This rule will be deleted anyway before achieving the time of hard.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide： msg.hard_timeout=30
Example：
Controller terminal:
Appoint the hard time of the rule as 30s.
POX> **msg.hard_timeout=30**
Explanation: None.

12.2.2 OPENFLOW Configuration (action)

Explanation: If there is no action in rules, it means to drop as default; the egress port needs to be added after the corresponding action if there is no display in the rules and the egress port was configured.

12.2.2.1 msg.actions.append(of.ofp_action_output(port=x))

Command：msg.actions.append(of.ofp_action_output(port=))
Function： Appoint the egress port action.
Parameters： None.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：msg.actions.append(of.ofp_action_output(port=))
Example：
Controller terminal: Appoint the egress port of the packet.
POX> **msg.actions.append(of.ofp_action_output(port=20))**
Explanation: The port number is the port in the "of vlan".

12.2.2.2 msg.actions.append(of.ofp_action_output(port=x))

Command：msg.actions.append(of.ofp_action_output(port=))
Function： Forward the appointed port type.
Parameters： IN_PORT = 0xfff8：send packets from the access port; FLOOD= 0xfffb：all the ports except the access ports and the ports which are not allowed by stp; ALL = 0xfffc：other ports except the access ports; CONTROLLER = 0xfffd： send to the controller; NONE = 0xffff：unrelated to the physical port.
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：msg.actions.append(of.ofp_action_output(port=))
Example：
Controller terminal:
POX> **msg.actions.append(of.ofp_action_output(port=all))**
Explanation: send a packet to all the ports except the access port.

12.2.2.3 msg.actions.append(of.ofp_action_enqueue(port=x，queue_id=y))

Command：msg.actions.append(of.ofp_action_enqueue(port=x,queue_id=y))
Function： Forward the appointed port and queue.
Parameters： port，queue_id (queue number)
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：msg.actions.append(of.ofp_action_enqueue(port=x，enqueue_id=y))
Example：
Controller terminal:
POX> **msg.actions.append(of.ofp_action_enqueue(port=13，queue_id=4))**

12.2.2.4 msg.actions.append(of.ofp_action_dl_addr.set_dst(“mac”))

Command：msg.actions.append(of.ofp_action_dl_addr.set_dst(“”))
Function： Change the destination mac to be the appointed mac.
Parameters： dst mac
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：msg.actions.append(of.ofp_action_dl_addr.set_dst(“11:11:11:11:11:11”))
Example：
Controller terminal:
POX> **msg.actions.append(of.ofp_action_dl_addr.set_dst(“11:11:11:11:11:11”))**
PE1(config-if-Vlan2)#ip address 120.1.1.2 255.255.255.

12.2.2.5 msg.actions.append(of.ofp_action_dl_addr.set_src(“mac”))

Command：msg.actions.append(of.ofp_action_dl_addr.set_src(“”))
Function： Change the source mac to be the appointed mac.
Parameters： src mac
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：msg.actions.append(of.ofp_action_dl_addr.set_src(“00:03:11:11:11:11”))
Example：
Controller terminal:
POX> **msg.actions.append(of.ofp_action_dl_addr.set_src(“00:03:11:11:11:11”))**
PE1(config-if-Vlan2)#ip address 120.1.1.2 255.255.255.

12.2.2.6 msg.actions.append(of.ofp_action_nw_tos(nw_tos=x))

Command：msg.actions.append(of.ofp_action_nw_tos(nw_tos=))
Function： Configure the tos value.
Parameters： nw_tos
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：msg.actions.append(of.ofp_action_nw_tos(nw_tos=56))
Example：
Controller terminal:
POX> **msg.actions.append(of.ofp_action_nw_tos(nw_tos=56))**

PE1(config-if-Vlan2)#ip address 120.1.1.2 255.255.255.

12.2.2.7 msg.actions.append(of.ofp_action_vlan_vid(vlan_vid=x))

Command：msg.actions.append(of.ofp_action_vlan_vid(vlan_vid=))
Function： Configure the vlan value.
Parameters： vlan_vid
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：msg.actions.append(of.ofp_action_vlan_vid(vlan_vid=3))
Example：
Controller terminal:
POX> **msg.actions.append(of.ofp_action_vlan_vid(vlan_vid=3))**

12.2.2.8 msg.actions.append(of.ofp_action_vlan_pcp(vlan_pcp=x))

Command：msg.actions.append(of.ofp_action_vlan_pcp(vlan_pcp=))
Function： Configure the vlan cos value.
Parameters： vlan cos
Default： None.
Command Mode： Interactive mode, pox>
Usage Guide：msg.actions.append(of.ofp_action_vlan_pcp(vlan_pcp=3))
Example：
Controller terminal:
POX> **msg.actions.append(of.ofp_action_vlan_vid(vlan_vid=3))**
POX> **msg.actions.append(of.ofp_action_vlan_pcp(vlan_pcp=4))**
Explanation: the vlan id must be configured first before configure the cos value.

12.2.3 OPENFLOW Configuration (examples)

12.2.3.1 add flow-match the access port, the action is the egress port

POX>import pox.openflow.libopenflow_01 as of
POX>msg2=of.ofp_flow_mod()
POX>msg2.priority=3
POX>msg2.match.in_port=193
POX>msg2.actions.append(of.ofp_action_output(port=194))

POX>core.openflow.connections[13136560386L].send(msg2)

12.2.3.2 add flow-match the destination mac, the action is the egress port

POX>import pox.openflow.libopenflow_01 as of
POX>msg2=of.ofp_flow_mod()
POX>msg2.priority=3
POX>msg2.match.dl_src=EthAddr("ff:ff:ff:ff:ff:ff")
POX>msg2.actions.append(of.ofp_action_output(port=194))
POX>core.openflow.connections[13136560386L].send(msg2)

12.2.3.3 add flow-match the type of Ethernet, the action is the egress port and queue

POX>msg=of.ofp_flow_mod()
POX>msg.priority=5
POX>msg.match.dl_type=0x800
POX>msg.actions.append(of.ofp_action_enqueue(queue_id=5,port=194))
POX>core.openflow.connections[13136560386L].send(msg)

12.2.3.4 add flow-match the source mac, the action is to configure the vlan and appoint the egress port

POX>msg=of.ofp_flow_mod()
POX>msg.priority=5
POX>msg.match.dl_src=EthAddr(“00:03:0f:01:12:43”)
POX>msg.actions.append(of.ofp_action_vlan_vid(vlan_vid=3))
POX>msg.actions.append(of.ofp_action_output(port=194))
POX>core.openflow.connections[13136560386L].send(msg)

12.2.3.5 add flow-match the access port, the action is to configure the vlan and cos and appoint the egress port

POX>msg=of.ofp_flow_mod()
POX>msg.priority=5
POX>msg.match.in_port=193
POX>msg.actions.append(of.ofp_action_vlan_vid(vlan_vid=4))
POX>msg.actions.append(of.ofp_action_vlan_pcp(vlan_pcp=5))
POX>msg.actions.append(of.ofp_action_output(port=194))

POX>core.openflow.connections[13136560386L].send(msg)

12.2.3.6 del flow

POX>msg=of.ofp_flow_mod(command=3)
POX>core.openflow.connections[13136560386L].send(msg)

12.2.3.7 del flow-strict

POX>msg=of.ofp_flow_mod(command=4)
POX>msg.wildcards= 4194302
POX>msg.priority=5
POX>core.openflow.connections[13136560386L].send(msg)

Chapter 13 vxlan Configuration

13.1 VXLAN

13.1.1 Introduction to VXLAN

VXLAN (Virtual Extensible Local Area Network) is a kind of technology which can package the layer 2 packets by the layer 3 protocol. It can extend the layer 2 network in the range of layer 3. It is applied to the data center interior and makes that the virtual machine can remove in the connected layer 3 networks without changing the IP address or MAC address and can ensure the continuity of the business. VXLAN adopts the 24bit network identity to make the user create 16M isolated virtual networks. It breaks through the limitation of the 4K isolated network with VLAN and this makes the cloud environment with massive multi-tenant have the enough virtual network partition resources.

VXLAN has the following characteristics:

Support plenty of tenants: it uses 24bit network identity; the supported number of tenants is increased. It solved the problem of insufficient resources of traditional layer 2 network.

Easy to maintain: it creates the larger layer 2 network based on IP network that makes the deployment and maintenance easier and it can make use of the existence IP network technology fully such as using the ECMP to conduct load-balance. Only the devices on the edge of the IP core network need the VXLAN configuration. The center devices only need to forward packets according to IP header. This reduces the difficulty and cost of the network deployment.

Save the switch MAC address table entries: the access switch only learns the MAC address of the physical server; it does not need to learn the MAC of every virtual machine. So it saves the switch performance of MAC table space promotion.

Currently, the device only supports the VXLAN based on IPv4 network. It does not support the VXLAN based on IPv6 network.

13.1.1.1 VXLAN Network Model

VXLAN network model is shown in the following figure, including:

VM (Virtual Machine): multiple VM can be created on one server. Different VM can belong to different VXLAN. The VMs belong to same VXLAN are in the same logic layer 2 network and they are communicating. The VMs belong to different VXLAN are isolated. VXLAN is marked by VXLAN ID. VXLAN ID is also named as VNI (VXLAN Network Identifier) and its range is 24bits.

VTEP (VXLAN Tunnel End Point): it is the edge device of VXLAN. The configuration of VXLAN is on VTEP, such as recognizing the VXLAN that the Ethernet data frame belongs to, conducting layer 2 forwarding for data frame based on VXLAN, packet encapsulation/de-encapsulation and so on. VTEP can be a independent physical device, and can also be the server that the VM is on.

VXLAN tunnel: it is the point-to-point logic tunnel between two VTEP. After VTEP encapsulated VXLAN header, UDP header and IP header for the data frame, the packet will be

sent to the terminal VETP through VXLAN tunnel and the terminal VTEP can de-encapsulate it

Core device: it is the device in the IP core network (such as P in the following figure). The core device does not deal with VXLAN, it only needs to conduct the layer 3 forwarding for the encapsulated packet according to the destination IP address of it.

NVI (Network Virtual Instance): it is the nvi which provides the layer 2 switch service for VXLAN on VTEP. NVI can be treated as a virtual switch which conducts layer 2 forwarding based on VXLAN on VTEP, and it has all the functions of the traditional Ethernet switch including MAC address learning, MAC address aging, flooding and so on. NVI and VXLAN are one-to-one correspondence.

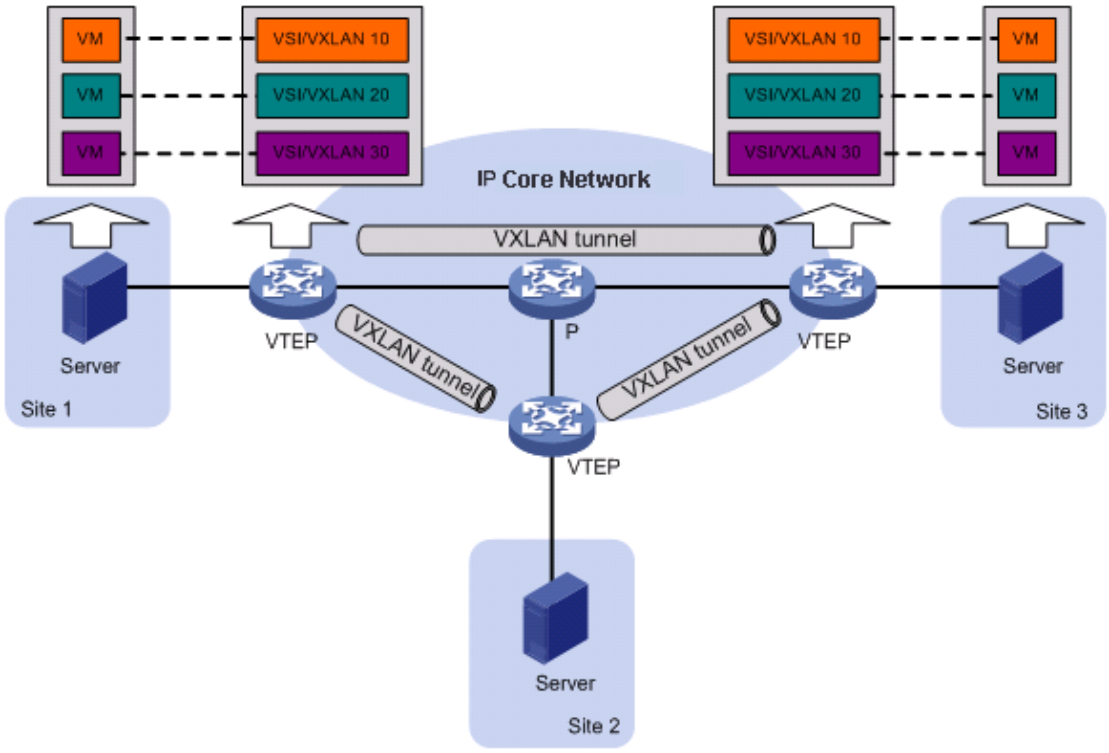


Figure 13-1 VXLAN network model

13.1.1.2 VXLAN Packet Packaging

The packaging format of VXLAN packet is: add 8 bytes VXLAN header, 8 bytes UDP header and 20 bytes IP header to the out of original layer 2 data frame. The destination port of UDP header is VXLAN UDP port (default value is 4789). The VXLAN header includes two parts:

Flag bit: when "I" is 1, it means that the VXLAN ID in VXLAN header is effective. When it is 0, it means that the VXLAN ID is not effective. Other bits are kept but not used and they are configured as 0.

VXLAN ID: it is a VXLAN network with 24 bits.

VXLAN packet packaging format is shown as below:

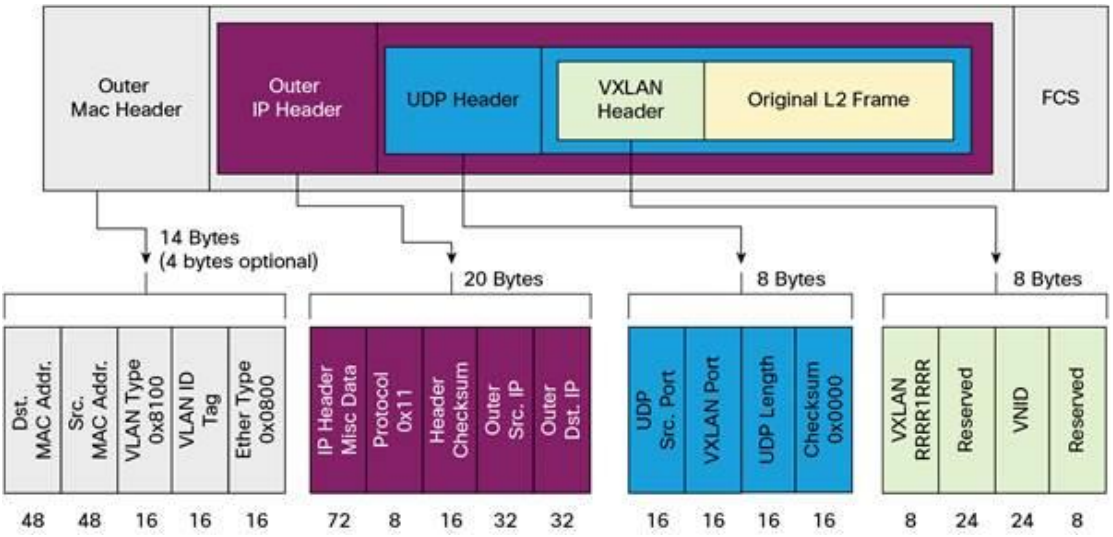


Figure 13-2 VXLAN packet packaging format

13.1.2 VXLAN Configuration

VXLAN configuration task list:

Configure nvi globally (Required)

Configure ID of vxlan under the nvi view (Required)

Configure nve tunnel interface globally (Required)

Configure vxlan access mapping on access port (Required)

View the nvi detailed information and received/sent packets statistics (Optional)

Configure nvi globally (Required)

Command	Explanation
Global Mode	
nvi <*nvi-id*> **no nvi <*nvi-id*>**	Create the network virtual instance or enter the network virtual instance view. The no command deletes the network virtual instance.

Configure ID of vxlan under the nvi view (Required)

Command	Explanation
NVI configuration Mode	
vxlan-id <*vxlan-id*>	Under the nvi view, configure the id of nvi and associate the nvi. The no command cancels the ID value.

Configure nve tunnel interface globally (Required)

Command	Explanation
Global Mode	

interface nve <*nve-id*> **no interface nve <*nve-id*>**	Create the vxlan tunnel interface or enter the vxlan tunnel view. The no command deletes the vxlan tunnel.
NVE Tunnel Interface Mode	
nve mode vxlan	Configure the tunnel mode.
source {tunnel-source-ip|word} **no source {tunnel-source-ip|word}**	Configure the ipv4 address of the source VTEP. The no command deletes it.
destination {tunnel-destination-ip|word} **no destination {tunnel-destination-ip|word}**	Configure the ipv4 address of the destination VTEP. The no command deletes it.
join nvi <*nvi-id* **no join nvi <*nvi-id*>**	Configure the association relationship between the nvi and vxlan tunnel. The no command cancels it.
description <*word*> **no description**	Appoint the nvi or nve tunnel interface name. the no command deletes it.

Configure vxlan access mapping on access port (Required)

Command	Explanation
Port Mode	
xconnect nvi <*nvi-id*> {|mode {ethernet|vlan svid<*vid*>}} **no xconnect nvi <*nvi-id*> {|mode {ethernet|vlan svid<*vid*>}}**	Configure the vxlan access mapping relationship of the access port.

View the nvi detailed information and received/sent packets statistics (Optional)

Command	Explanation
Any Mode	
show interface nve <*nve-id*>	View the received and sent packets and bytes statistics of nve tunnel interface.
show nvi <*nvi-id*> detail	Show the detailed information of nvi.

13.1.3 VXLAN Example

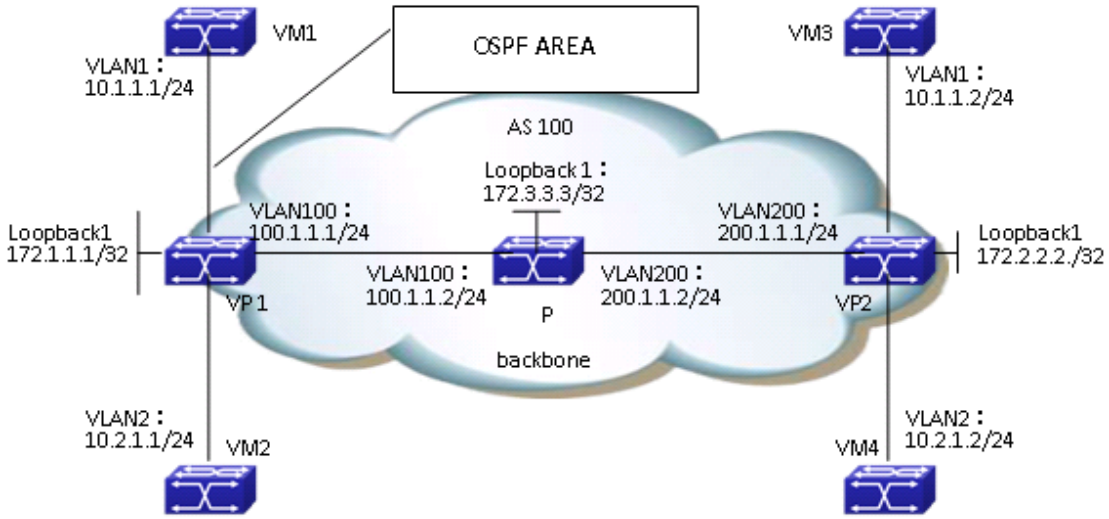


Figure 13-3 VXLAN typical example

Configuration steps:

Please configure the IP address and mask for each interface according to the above figure, including VLAN interface and loopback interface. The detailed configuration will not be shown here, there is only the VXLAN configuration.

VP1 configuration:

#Configure nvi
VP1(config)#nvi 1
VP1(config-nvi)#vxlan-id 1
VP1(config-nvi)#description VM1
VP1(config)#nvi 2
VP1(config-nvi)#vxlan-id 2
VP1(config-nvi)#description VM2
#Configure vxlan tunnel interface
VP1(config)#interface nve 1
VP1(config-if-nve1)#nve mode vxlan
VP1(config-if-nve1)#source 172.1.1.1
VP1(config-if-nve1)#destination 172.2.2.2
VP1(config-if-nve1)#join nvi 1
VP1(config-if-nve1)#join nvi 2
VP1(config-if-nve1)#description VP1
#Configure the vxlan mapping of access interface
VP1(config)#interface ethernet 1/0/1
VP1(config-if-ethernet1/0/1)#switchport mode access
VP1(config-if-ethernet1/0/1)#description VM1
VP1(config-if-ethernet1/0/1)#xconnect nvi 1 mode vlan svid 1
VP1(config)#interface ethernet 1/0/2
VP1(config-if-ethernet1/0/2)#switchport mode access
VP1(config-if-ethernet1/0/1)#description VM2
VP1(config-if-ethernet1/0/2)#xconnect nvi 2 mode vlan svid 2

VP2 configuration:

```
#Configure nvi
VP2(config)#nvi 1
VP2(config-nvi)#vxlan-id 1
VP2(config-nvi)#description VM3
VP2(config)#nvi 2
VP2(config-nvi)#vxlan-id 2
VP2(config-nvi)#description VM4
# Configure vxlan tunnel interface
VP2(config)#interface nve 1
VP2(config-if-nve1)#nve mode vxlan
VP2(config-if-nve1)#source 172.2.2.2
VP2(config-if-nve1)#destination 172.1.1.1
VP2(config-if-nve1)#join nvi 1
VP2(config-if-nve1)#join nvi 2
VP2(config-if-nve1)#description VP2
# Configure the vxlan mapping of access interface
VP2(config)#interface ethernet 1/0
VP2(config-if-ethernet1/0/1)#switchport mode access
VP2(config-if-ethernet1/0/1)#description VM3
VP2(config-if-ethernet1/0/1)#xconnect nvi 1 mode vlan svid 1
VP2(config)#interface ethernet 1/0/2
VP2(config-if-ethernet1/0/2)#switchport mode access
VP2(config-if-ethernet1/0/1)#description VM4
VP2(config-if-ethernet1/0/2)#xconnect nvi 2 mode vlan svid 2
```

Through configuring the vxlan mapping of access interface, VM1 and VM3 can communicate, and VM2 and VM4 can communicate. VM1 and VM2, VM3 and VM4 are isolated respectively.

13.1.4 VXLAN Troubleshooting

When configuring and using VXLAN, it may not run normally because of the physical connection or wrong configuration. User should pay attention to the following points:

- Firstly, please ensure the OSPF neighbor among VETP1, P and VETP2 is created normally, including router issued normally.
- Then, ensure the VXLAN tunnel between VETP1 and VETP2 is configured correctly, and the source and destination routers are reachable.
- And then, ensure the access interface is configured correctly. The access vxlan mapping includes Ethernet mode and vlan mode, the edge access interfaces are configured conformably.
- Last, view the achieved and sent packets statistics of nvi and nve interface through the show command for error checking.

Chapter 14 IPv6 Configuration

14.1 DHCPv6

14.1.1 Introduction to DHCPv6

DHCPv6 [RFC3315] is the IPv6 version for Dynamic Host Configuration Protocol (DHCP). It is a protocol that assigns IPv6 address as well as other network configuration parameters such as DNS address, and domain name to DHCPv6 client, DHCPv6 is a conditional auto address configuration protocol relative to IPv6. In the conditional address configuration process, DHCPv6 server assigns a complete IPv6 address to client, and provides DNS address, domain name and other configuration information, maybe the DHCPv6 packet can transmit through relay delegation, at last the binding of IPv6 address and client can be recorded by DHCPv6 server, all that can enhance the management of network; DHCPv6 server can also provide non state DHCPv6 service, that is only assigns DNS address and domain name and other configuration information but not assigns IPv6 address, it can solve the bug of IPv6 auto address configuration in non state; DHCPv6 can provide extend function of DHCPv6 prefix delegation, upstream route can assign address prefix to downstream route automatically, that achieve the IPv6 address auto assignment in levels of network environment, and resolved the problem of ISP and IPv6 network dispose.

There are three entities in the DHCPv6 protocol – the client, the relay and the server. The DHCPv6 protocol is based on the UDP protocol. The DHCPv6 client sends request messages to the DHCP server or DHCP relay with the destination port as 547, and the DHCPv6 server and relay send replying messages with the destination port as 546. The DHCPv6 client sends solicit or request messages with the multicast address – ff02::1:2 for DHCP relay and server.

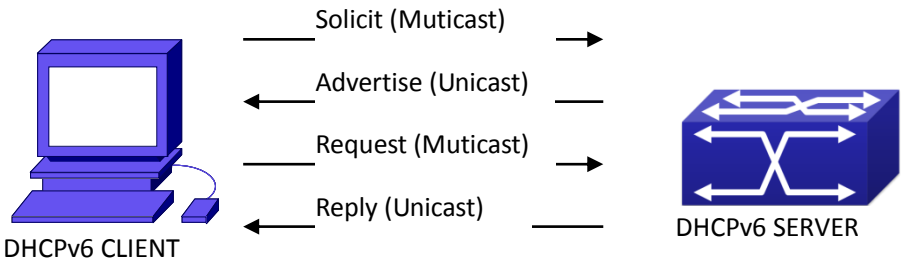


Figure 14-1 DHCPv6 negotiation

When a DHCPv6 client tries to request an IPv6 address and other configurations from the DHCPv6 server, the client has to find the location of the DHCP server, and then request configurations from the DHCP server.

1. In the time of located server, the DHCP client tries to find a DHCPv6 server by broadcasting a SOLICIT packet to all the DHCP delay delegation and server with broadcast address as FF02::1:2.

2. Any DHCP server which receives the request, will reply the client with an ADVERTISE message, which includes the identity of the server –DUID, and its priority.
3. It is possible that the client receives multiple ADVERTISE messages. The client should select one and reply it with a REQUEST message to request the address which is advertised in the ADVERTISE message.
4. The selected DHCPv6 server then confirms the client about the IPv6 address and any other configuration with the REPLY message.

The above four steps finish a Dynamic host configuration assignment process. However, if the DHCPv6 server and the DHCPv6 client are not in the same network, the server will not receive the DHCPv6 broadcast packets sent by the client, therefore no DHCPv6 packets will be sent to the client by the server. In this case, a DHCPv6 relay is required to forward such DHCPv6 packets so that the DHCPv6 packets exchange can be completed between the DHCPv6 client and server.

At the time this manual is written, DHCPv6 server, relay and prefix delegation client have been implemented on the switch. When the DHCPv6 relay receives any messages from the DHCPv6 client, it will encapsulate the request in a Relay-forward packet and deliver it to the next DHCPv6 relay or the DHCPv6 server. The DHCPv6 messages coming from the server will be encapsulated as relay reply packets to the DHCPv6 relay. The relay then removes the encapsulation and delivers it the DHCPv6 client or the next DHCPv6 relay in the network.

For DHCPv6 prefix delegation where DHCPv6 server is configured on the PE router and DHCPv6 client it configured on the CPE router, the CPE router is able to send address prefix allocation request to the PE router and get a pre-configured address prefix, but not set the address prefix manually. The protocol negotiation between the client and the prefix delegation client is quite similar to that when getting a DHCPv6 address. Then the CPE router divides the allocated prefix – whose length should be less than 64 characters, into 64 subnets. The divided address prefix will be advertised through routing advertisement messages (RA) to the host directly connected to the client.

14.1.2 DHCPv6 Server Configuration

DHCPv6 server configuration task list as below:

1. To enable/disable DHCPv6 service
2. To configure DHCPv6 address pool
   （1） To achieve/delete DHCPv6 address pool
   （2） To configure parameter of DHCPv6 address pool
3. To enable DHCPv6 server function on port

1. To enable/disable DHCPv6 service

Command	Explanation
Global Mode	

service dhcpv6 **no service dhcpv6**	To enable DHCPv6 service.

2. To configure DHCPv6 address pool

（1）To achieve/delete DHCPv6 address pool

Command	Explanation
Global Mode	
ipv6 dhcp pool <*poolname*> **no ipv6 dhcp pool <*poolname*>**	To configure DHCPv6 address pool.

（2）To configure parameter of DHCPv6 address pool

Command	Explanation
DHCPv6 address pool Configuration Mode	
network-address <*ipv6-pool-start-address*> {<*ipv6-pool-end-address*> | <*prefix-length*>} [eui-64] **no network-address**	To configure the range of IPv6 address assignable of address pool.
dns-server <*ipv6-address*> **no dns-server <*ipv6-address*>**	To configure DNS server address for DHCPv6 client.
domain-name <*domain-name*> **no domain-name <*domain-name*>**	To configure DHCPv6 client domain name.
excluded-address <*ipv6-address*> **no excluded-address <*ipv6-address*>**	To exclude IPv6 address which isn't used for dynamic assignment in address pool.
lifetime {<*valid-time*> | infinity} {<*preferred-time*> | infinity} **no lifetime**	To configure valid time or preferred time of DHCPv6 address pool.

3. To enable DHCPv6 server function on port.

Command	Explanation
Interface Configuration Mode	
ipv6 dhcp server <*poolname*> [preference <*value*>] [rapid-commit] [allow-hint] **no ipv6 dhcp server <*poolname*>**	To enable DHCPv6 server function on specified port, and binding the used DHCPv6 address pool.

14.1.3 DHCPv6 Relay Delegation Configuration

DHCPv6 relay delegation configuration task list as below:

1. To enable/disable DHCPv6 service
2. To configure DHCPv6 relay delegation on port

1. To enable DHCPv6 service

Command	Explanation

Global Mode	
service dhcpv6 **no service dhcpv6**	To enableDHCPv6 service.

2. To configure DHCPv6 relay delegation on port

Command	Explanation
Interface Configuration Mode	
ipv6 dhcp relay destination {[<*ipv6-address*>] [interface { <*interface-name*> | vlan <*1-4096*>}]} **no ipv6 dhcp relay destination {[<*ipv6-address*>] [interface { <*interface-name*> | vlan <*1-4096*>}]}**	To specify the destination address of DHCPv6 relay transmit; The no form of this command delete the configuration.

14.1.4 DHCPv6 Prefix Delegation Server Configuration

DHCPv6 prefix delegation server configuration task list as below:

1. To enable/delete DHCPv6 service
2. To configure prefix delegation pool
3. To configure DHCPv6 address pool
   （1） To achieve/delete DHCPv6 address pool
   （2） To configure prefix delegation pool used by DHCPv6 address pool
   （3） To configure static prefix delegation binding
   （4） To configure other parameters of DHCPv6 address pool
4. To enable DHCPv6 prefix delegation server function on port

1. To enable/delete DHCPv6 service

Command	Explanation
Global Mode	
service dhcpv6 **no service dhcpv6**	To enable DHCPv6 service.

2. To configure prefix delegation pool

Command	Explanation
Global Mode	
ipv6 local pool <*poolname*> <*prefix*|*prefix-length*> <*assigned-length*> **no ipv6 local pool <*poolname*>**	To configure prefix delegation pool.

3. To configure DHCPv6 address pool

（1）To achieve/delete DHCPv6 address pool

Command	Explanation
Global Mode	

ipv6 dhcp pool <*poolname*> **no ipv6 dhcp pool <*poolname*>**	To configure DHCPv6 address pool.

（2）To configure prefix delegation pool used by DHCPv6 address pool

Command	Explanation
DHCPv6 address pool Configuration Mode	
prefix-delegation pool <*poolname*> [lifetime {<*valid-time*> / *infinity*} {<*preferred-time*> | infinity}] **no *prefix-delegation pool* <poolname>**	To specify prefix delegation pool used by DHCPv6 address pool, and assign usable prefix to client.

（3） To configure static prefix delegation binding

Command	Explanation
DHCPv6 address pool Configuration Mode	
prefix-delegation <*ipv6-prefix/prefix-length*> <*client-DUID*> [iaid <*iaid*>] [lifetime {<valid-time> | infinity} {<*preferred-time*> | infinity}] **no *prefix-delegation* <*ipv6-prefix*/prefix-length> <client-DUID> [iaid <*iaid*>]**	To specify IPv6 prefix and any prefix required static binding by client.

（4） To configure other parameter of DHCPv6 address pool

Command	Explanation
DHCPv6 address pool Configuration Mode	
dns-server <*ipv6-address*> **no dns-server <*ipv6-address*>**	To configure DNS server address for DHCPv6 client.
domain-name <*domain-name*> **no domain-name <*domain-name*>**	To configure domain name for DHCPv6 client.

4. To enable DHCPv6 prefix delegation server function on port

Command	Explanation
Interface Configuration Mode	
ipv6 dhcp server <*poolname*> [preference <*value*>] [rapid-commit] [allow-hint] **no ipv6 dhcp server <*poolname*>**	To enable DHCPv6 server function on specified port, and binding used DHCPv6 address pool.

14.1.5 DHCPv6 Prefix Delegation Client Configuration

DHCPv6 prefix delegation client configuration task list as below:

1. To enable/disable DHCPv6 service
2. To enable DHCPv6 prefix delegation client function on port

1. To enable/disable DHCPv6 service

Command	Explanation
Global Mode	
service dhcpv6 **no service dhcpv6**	To enable DHCPv6 service.

2. To enable DHCPv6 prefix delegation client function on port

Command	Explanation
Interface Configuration Mode	
ipv6 dhcp client pd <*prefix-name*> [rapid-commit] **no ipv6 dhcp client pd**	To enable client prefix delegation request function on specified port, and the prefix obtained associate with universal prefix configured.

14.1.6 DHCPv6 Configuration Examples

Example1:

When deploying IPv6 networking, the switch can be configured as DHCPv6 server in order to manage the allocation of IPv6 addresses. Both the state and the stateless DHCPv6 are supported.

Topology:

The access layer use Switch1 switch to connect users of dormitory buildings; Switch2 is configured as DHCPv6 relay delegation in primary aggregation layer ; Switch3 is configured as DHCPv6 server in secondary aggregation layer, and connected with backbone network or higher aggregation layers; The Windows Vista which be provided with DHCPv6 client must load on PC.

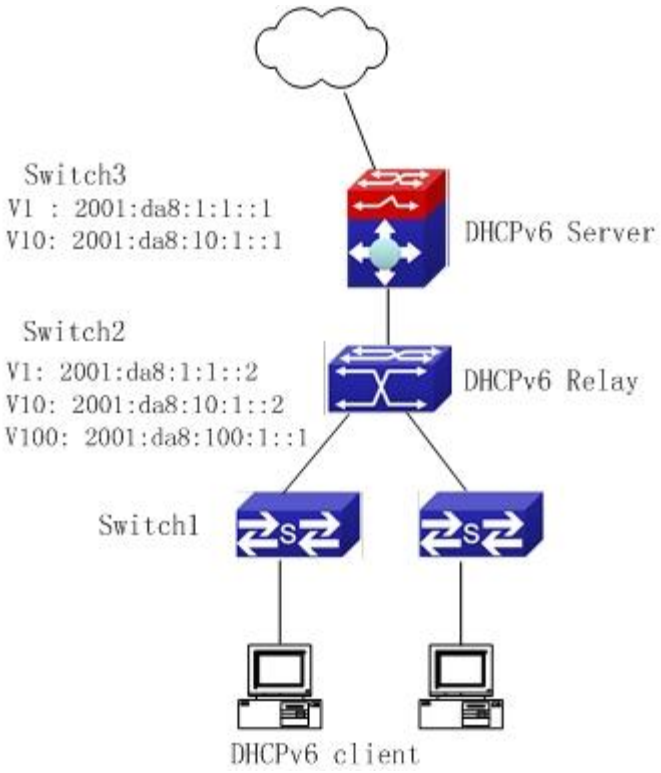


Figure 14-2

Usage guide:

Switch3 configuration：

```
Switch3>enable
Switch3#config
Switch3(config)#service dhcpv6
Switch3(config)#ipv6 dhcp pool EastDormPool
Switch3(dhcpv6-EastDormPool-config)#network-address 2001:da8:100:1::1 2001:da8:100:1::100
Switch3(dhcpv6-EastDormPool-config)#excluded-address 2001:da8:100:1::1
Switch3(dhcpv6-EastDormPool-config)#dns-server 2001:da8::20
Switch3(dhcpv6-EastDormPool-config)#dns-server 2001:da8::21
Switch3(dhcpv6-EastDormPool-config)#domain-name dhcpv6.com
Switch3(dhcpv6-EastDormPool-config)#lifetime 1000 600
Switch3(dhcpv6-EastDormPool-config)#exit
Switch3(config)#interface vlan 1
Switch3(Config-if-Vlan1)#ipv6 address 2001:da8:1:1::1/64
Switch3(Config-if-Vlan1)#exit
Switch3(config)#interface vlan 10
Switch3(Config-if-Vlan10)#ipv6 address 2001:da8:10:1::1/64
Switch3(Config-if-Vlan10)#ipv6 dhcp server EastDormPool preference 80
Switch3(Config-if-Vlan10)#exit
Switch3(config)#
```

Switch2 configuration：

```
Switch2>enable
Switch2#config
Switch2(config)#service dhcpv6
Switch2(config)#interface vlan 1
Switch2(Config-if-Vlan1)#ipv6 address 2001:da8:1:1::2/64
Switch2(Config-if-Vlan1)#exit
Switch2(config)#interface vlan 10
Switch2(Config-if-Vlan10)#ipv6 address 2001:da8:10:1::2/64
Switch2(Config-if-Vlan10)#exit
Switch2(config)#interface vlan 100
Switch2(Config-if-Vlan100)#ipv6 address 2001:da8:100:1::1/64
Switch2(Config-if-Vlan100)#no ipv6 nd suppress-ra
Switch2(Config-if-Vlan100)#ipv6 nd managed-config-flag
Switch2(Config-if-Vlan100)#ipv6 nd other-config-flag
Switch2(Config-if-Vlan100)#ipv6 dhcp relay destination 2001:da8:10:1::1
Switch2(Config-if-Vlan100)#exit
Switch2(config)#
```

Example2:

When the network operator is deploying IPv6 networks, network automatically configuration can be achieved through the prefix delegation allocation of IPv6 addresses, in stead of configuring

manually for each switch:

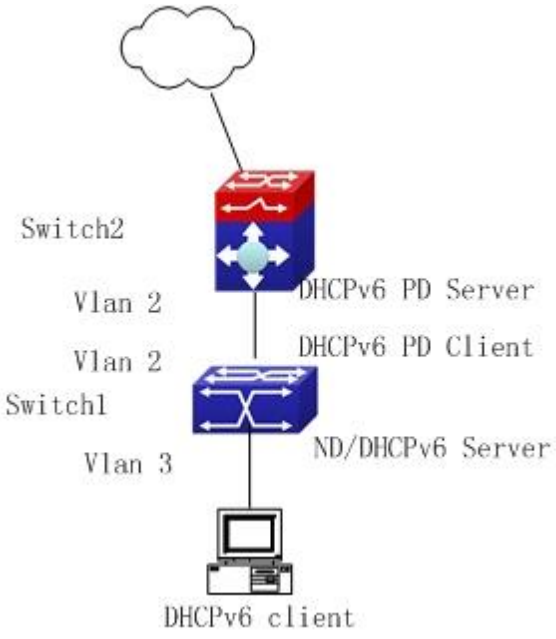
To configure the switching or routing device which is connected to the client switch as DHCPv6 prefix delegation server, that is to setup a local database for the relationship between the allocated prefix and the DUID of the client switch.

To configure the switch as the prefix delegation client, and make the client switch to get IPv6 address prefix from the prefix delegation server, through a process which is much like the process of DHCPv6 address allocation.

The edge devices which receive the address prefix, send routing advertisement - RA messages, to the client hosts about the address prefix through the interface which is connected to the hosts, then the hosts get an valid IPv6 address through stateless auto configuration, while at the same time, the stateless DHCPv6 server will be configured for the interface, in order to provide the DHCPv6 client with information such as DNS, and domain name, etc.

Network Topology:

The edge switch is a Switch1 switch. The interface connected to the trunk switch which is Switch2, is configured as the prefix delegation client. The interfaces connected to hosts, are configured as stateless DHCPv6 servers to provide the hosts with stateless information such as DNS and domain names, also routing advertisement of stateless address allocation is enabled for the host interfaces; On Switch2, the prefix delegation server is configured, and routing advertisement of state address allocation is enabled; On the host side, DHCPv6 client capable operating system such Windows Vista should be installed.



Usage guide:

Switch2 configuration

Switch2>enable

Switch2#config

Switch2(config)#interface vlan 2

Switch2(Config-if-Vlan2)#ipv6 address 2001:da8:1100::1/64

Switch2(Config-if-Vlan2)#exit

Switch2(config)#service dhcpv6

```
Switch2(config)#ipv6 local pool client-prefix-pool 2001:da8:1800::/40 48
Switch2(config)#ipv6 dhcp pool dhcp-pool
Switch2(dhcpv6-dhcp-pool-config)#prefix-delegation pool client-prefix-pool 1800 600
Switch2(dhcpv6-dhcp-pool-config)#exit
Switch2(config)#interface vlan 2
Switch2(Config-if-Vlan2)#ipv6 dhcp server dhcp-pool
Switch2(Config-if-Vlan2)#exit

Switch1 configuration
Switch1>enable
Switch1#config
Switch1(config)#service dhcpv6
Switch1(config)#interface vlan 2
Switch1(Config-if-Vlan2)#ipv6 dhcp client pd prefix-from-provider
Switch1(Config-if-Vlan2)#exit
Switch1(config)#interface vlan 3
Switch1(Config-if-Vlan3)#ipv6 address prefix-from-provider 0:0:0:1::1/64
Switch1(Config-if-Vlan3)#exit
Switch1(config)#ipv6 dhcp pool foo
Switch1(dhcpv6-foo-config)#dns-server 2001:4::1
Switch1(dhcpv6-foo-config)#domain-name www.ipv6.org
Switch1(dhcpv6-foo-config)#exit
Switch1(config)#interface vlan 3
Switch1(Config-if-Vlan3)#ipv6 dhcp server foo
Switch1(Config-if-Vlan3)#ipv6 nd other-config-flag
Switch1(Config-if-Vlan3)#no ipv6 nd suppress-ra
Switch1(Config-if-Vlan3)#exit
```

14.1.7 DHCPv6 Troubleshooting

If the DHCPv6 clients cannot obtain IPv6 addresses and other network parameters, the following procedures can be followed when DHCPv6 client hardware and cables have been verified ok:

- ☞ Verify the DHCPv6 server is running, start the related DHCP v6 server function if not running;
- ☞ If the DHCPv6 clients and servers are not in the same physical network, verify the router responsible for DHCPv6 packet forwarding has DHCPv6 relay function. If DHCPv6 relay is not available for the intermediate router, it is recommended to replace the router or upgrade its software to one that has a DHCPv6 relay function;
- ☞ Sometimes hosts are connected to the DHCPv6 enabled switches, but can not get IPv6 addresses. In this situation, it should be checked first whether the ports which the hosts are connected to, are connected with the port which the DHCPv6 server is connected to. If

connected directly, it should be checked then whether the IPv6 address pool of the VLAN which the port belongs to, is in the same subnet with the address pool configure in the DHCPv6 server; If not connected directly, and any layer three DHCPv6 relay is configured between the hosts and the DHCPv6 server, it should be checked first whether an valid IPv6 address has been configured for the switch interface which the hosts are connected to. If not configured, configure an valid IPv6 address. If configured, it should be checked whether the configured IPv6 address is in the same subnet with the DHCPv6 server. If not, please add it to the address pool.

14.2 DHCPv6 option37, 38

14.2.1 Introduction to DHCPv6 option37, 38

DHCPv6 (Dynamic Host Configuration Protocol for IPv6) is designed for IPv6 address scheme and is used for assigning IPv6 prefixes, IPv6 addresses and other configuration parameters to hosts.

When DHCPv6 client wants to request address and configure parameter of DHCPv6 server from different link, it needs to communicate with server through DHCPv6 relay agent. DHCPv6 message received by relay agent node is reencapsulated to be relay-forward packets and they are forwarded to the server which sends the relay-reply packets to DHCPv6 relay agent node in different link, after that, relay agent node restores DHCPv6 message to DHCPv6 client to finish communication between client and server.

There are some problems when using DHCPv6 relay agent, for example: How to assign IP address in the fixed range to the specifiec users? How to avoid illegal DHCPv6 client to forge IP address exhaust attack triggered by MAC address fields of DHCPv6 packets? How to avoid illegal DHCPv6 client to trigger deny service attack through using MAC address of other legal clients? Therefore, IETF set rfc4649 and rfc4580, i.e. DHCPv6 option 37 and option 38 to solve these problems.

DHCPv6 option 37 and option 38 is similar to DHCP option 82. When DHCPv6 client sends request packets to DHCPv6 server though DHCPv6 relay agent, if DHCPv6 relay agent supports option 37 and option 38, they will be added to request packets. For the respond packets of server, option 37 and option 38 are meaningless and are peeled from the respond packets. Therefore, the application of option 37 and option 38 is transparent for client.

DHCPv6 server can authenticate identity of DHCPv6 client and DHCPv6 relay device by option 37 and option 38, assign and manage client address neatly through configuring the assign policy, prevent DHCPv6 attack availably according to the inclusive client information, such as forging MAC address fields of DHCPv6 packets to trigger IP address exhaust attack. Since server can identify multiple request packets from the same access port, it can assign the address number through policy limit to avoid address exhaust. However, rfc4649 and rfc4580 do not set how to use opton 37 and option 38 for DHCPv6 server, users can use it neatly according to their

own demand.

14.2.2 DHCPv6 option37, 38 Configuration Task List

1. Dhcpv6 snooping option basic functions configuration
2. Dhcpv6 relay option basic functions configuration
3. Dhcpv6 server option basic functions configuration

1.DHCPv6 snooping option basic functions configuration

Command	Description
Global mode	
ipv6 dhcp snooping remote-id option **no ipv6 dhcp snooping remote-id option**	This command enables DHCPv6 SNOOPING to support option 37 option, no command disables it.
ipv6 dhcp snooping subscriber-id option **no ipv6 dhcp snooping subscriber-id option**	This command enables DHCPv6 SNOOPING to support option 38 option, no command disables it.
ipv6 dhcp snooping remote-id policy {drop | keep | replace} **no ipv6 dhcp snooping remote-id policy**	This command is used to configure the reforward policy of the system when receiving DHCPv6 packets with option 37, which can be: **drop,** the system simply discards it with option 37; **keep,** the system keeps option 37 unchanged and forwards the packet to the server; **replace,** the system replaces option 37 of current packet with its own before forwarding it to the server. no command configures the reforward policy of DHCPv6 packets with option 37 as replace.
ipv6 dhcp snooping subscriber-id policy {drop | keep | replace} **no ipv6 dhcp snooping subscriber-id policy**	This command is used to configure the reforward policy of the system when receiving DHCPv6 packets with option 38, which can be: **drop,** the system simply discards it with option 38; **keep,** the system keeps option

	38 unchanged and forwards the packet to the server; **replace,** the system replaces option 38 of current packet with its own before forwarding it to the server. no command configures the reforward policy of DHCPv6 packets with option 38 as replace.
ipv6 dhcp snooping subscriber-id select (sp | sv | pv | spv) delimiter WORD (delimiter WORD |) **no ipv6 dhcp snooping subscriber-id select delimiter**	Configures user configuration options to generate subscriber-id, no command restores to its original default configuration, i.e. enterprise number together with vlan MAC.
ipv6 dhcp snooping subscriber-id select (sp|sv|pv|spv) delimiter WORD (delimiter WORD |) **no ipv6 dhcp snooping subscriber-id select delimiter**	Configures user configuration options to generate subscriber-id. The no command restores to its original default configuration, i.e. vlan name together with port name.
Port mode	
ipv6 dhcp snooping remote-id <remote-id> **no ipv6 dhcp snooping remote-id**	This command is used to set the form of adding option 37 in received DHCPv6 request packets, of which <remote-id> is the content of remote-id in user-defined option 37 and it is a string with a length of less than 128. The no operation restores remote-id in option 37 to enterprise-number together with vlan MAC address.
ipv6 dhcp snooping subscriber-id <subscriber-id> **no ipv6 dhcp snooping subscriber-id**	This command is used to set the form of adding option 38 in received DHCPv6 request packets, of which <subscriber-id> is the content of subscriber-id in user-defined option 38 and it is a string with a length of less than 128. The no operation restores subscriber-id in option 38 to

	vlan name together with port name such as "Vlan2+Ethernet1/0/2".

2. DHCPv6 relay option basic functions configuration

Command	Description
Global mode	
ipv6 dhcp relay remote-id option **no ipv6 dhcp relay remote-id option**	This command enables the switch relay to support option 37 and the no form of this command disables it.
ipv6 dhcp relay subscriber-id option **no ipv6 dhcp relay subscriber-id option**	This command enables the switch relay to support the option 38, the no form of this command disables it.
ipv6 dhcp relay remote-id delimiter WORD **no ipv6 dhcp relay remote-id delimiter**	Configures user configuration options to generate remote-id. The no command restores to its original default configuration, i.e. enterprise number together with vlan MAC.
ipv6 dhcp relay subscriber-id select (sp | sv | pv | spv) delimiter WORD (delimiter WORD |) **no ipv6 dhcp relay subscriber-id select delimiter**	Configures user configuration options to generate subscriber-id. The no command restores to its original default configuration, i.e. vlan name together with port name.
Layer3 Interface configuration mode	
ipv6 dhcp relay remote-id <remote-id> **no ipv6 dhcp relay remote-id**	This command is used to set the form of adding option 37 in received DHCPv6 request packets, of which <remote-id> is the content of remote-id in user-defined option 37 and it is a string with a length of less than 128. The no operation restores remote-id in option 37 to enterprise-number together with vlan MAC address.
ipv6 dhcp relay subscriber-id <subscriber-id> **no ipv6 dhcp relay subscriber-id**	This command is used to set the form of adding option 38 in received DHCPv6 request packets, of which <subscriber-id> is the content

	of subscriber-id in user-defined option 38 and it is a string with a length of less than 128. The no operation restores subscriber-id in option 38 to vlan name together with port name such as "Vlan2+Ethernet1/0/2".

3. Dhcpv6 server option basic functions configuration

Command	Description
Global mode	
ipv6 dhcp server remote-id option **no ipv6 dhcp server remote-id option**	This command enables DHCPv6 server to support the identification of option 37, the no form of this command disables it.
ipv6 dhcp server subscriber-id option **no ipv6 dhcp server subscriber-id option**	This command enables DHCPv6 server to support the identification of option 38, the no form of this command disables it.
ipv6 dhcp use class **no ipv6 dhcp use class**	This command enables DHCPv6 server to support the using of DHCPv6 class during address assignment, the no form of this command disables it without removing the relative DHCPv6 class information that has been configured.
ipv6 dhcp class <class-name> **no ipv6 dhcp class <class-name>**	This command defines a DHCPv6 class and enters DHCPv6 class mode, the no form of this command removes this DHCPv6 class.
Interface configuration mode	
ipv6 dhcp server select relay-forw **no ipv6 dhcp server select relay-forw**	This command enables the DHCPv6 server to support selections when multiple option 37 or option 38 options exist and the option 37 and option 38 of relay-forw in the innermost layer are selected. The no operation of it restores

	the default configuration, i.e. selecting option 37 and option 38 of the original packets.
IPv6 DHCP Class configuration mode	
{remote-id [*] <remote-id> [*] | subscriber-id [*] <subscriber-id> [*]} **no {remote-id [*] <remote-id> [*] | subscriber-id [*] <subscriber-id> [*]}**	This command configures option 37 and option 38 that match the class in ipv6 dhcp class configuration mode.
DHCPv6 address pool configuration mode	
class <class-name> **no class <class-name>**	This command associates class to address pool in DHCPv6 address pool configuration mode and enters class configuration mode in address pool. Use no command to remove the link.
address range <start-ip> <end-ip> **no address range <start-ip> <end-ip>**	This command is used to set address range for a DHCPv6 class in DHCPv6 address pool configuration mode, the no command is used to remove the addreass range. The prefix/plen form is not supported.

14.2.3 DHCPv6 option37, 38 Examples

14.2.3.1 DHCPv6 Snooping option37, 38 Example

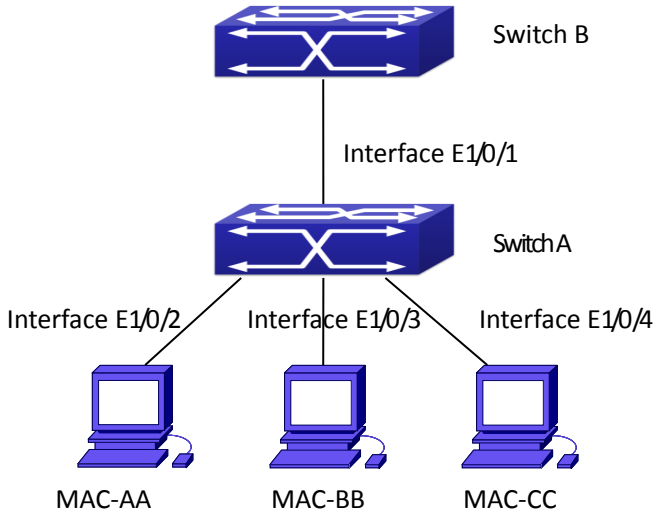


Figure 14-4 DHCPv6 Snooping option schematic

As is shown in the figure above, Mac-AA, Mac-BB and Mac-CC are normal users, connected to untrusted interface 1/0/2, 1/0/3 and 1/0/4 respectively, and they get IP 2010:2, 2010:3 and 2010:4 through DHCPv6 Client; DHCPv6 Server is connected to the trusted interface 1/0/1. Configure three address assignment policies (CLASS), of which CLASS1 matches option 38, CLASS2 matches option 37 and CLASS3 matches option 37 and option 38. In the address pool EastDormPool, the requests matched with CLASS1, CLASS2 and CLASS3 will be assigned an address ranging from 2001:da8:100:1::2 to 2001:da8:100:1::30, from 2001:da8:100:1::31 to 2001:da8:100:1::60 and from 2001:da8:100:1::61 to2001:da8:100:1::100 respectively; DHCPv6 snooping function is enabled and option 37 and option 38 are configured in Switch A.

Switch A configuration:

```
SwitchA(config)#ipv6 dhcp snooping remote-id option
SwitchA(config)#ipv6 dhcp snooping subscriber-id option
SwitchA(config)#int e 1/0/1
SwitchA(config-if-ethernet1/0/1)#ipv6 dhcp snooping trust
SwitchA(config-if-ethernet1/0/1)#exit
SwitchA(config)#interface vlan 1

SwitchA(config-if-vlan1)#ipv6 address 2001:da8:100:1::1
SwitchA(config-if-vlan1)#exit
SwitchA(config)#interface ethernet 1/0/1-4
SwitchA(config-if-port-range)#switchport access vlan 1
SwitchA(config-if-port-range)#exit
SwitchA(config)#
```

Switch B configuration:

```
SwitchB(config)#service dhcpv6
SwitchB(config)#ipv6 dhcp server remote-id option
```

```
SwitchB(config)#ipv6 dhcp server subscriber-id option
SwitchB(config)#ipv6 dhcp pool EastDormPool
SwitchB(dhcpv6-eastdormpool-config)#network-address 2001:da8:100:1::2
2001:da8:100:1::1000
SwitchB(dhcpv6-eastdormpool-config)#dns-server 2001::1
SwitchB(dhcpv6-eastdormpool-config)#domain-name dhcpv6.com
SwitchB(dhcpv6-eastdormpool-config)# excluded-address 2001:da8:100:1::2
SwitchB(dhcpv6-eastdormpool-config)#exit
SwitchB(config)#
SwitchB(config)#ipv6 dhcp class CLASS1
SwitchB(dhcpv6-class-class1-config)#remote-id 00-03-0f-00-00-01 subscriber-id
vlan1+Ethernet1/0/1
SwitchB(dhcpv6-class-class1-config)#exit
SwitchB(config)#ipv6 dhcp class CLASS2
SwitchB(dhcpv6-class-class2-config)#remote-id 00-03-0f-00-00-01 subscriber-id
vlan1+Ethernet1/0/2
SwitchB(dhcpv6-class-class2-config)#exit
SwitchB(config)#ipv6 dhcp class CLASS3
SwitchB(dhcpv6-class-class3-config)#remote-id 00-03-0f-00-00-01 subscriber-id
vlan1+Ethernet1/0/3
SwitchB(dhcpv6-class-class3-config)#exit
SwitchB(config)#ipv6 dhcp pool EastDormPool
SwitchB(dhcpv6-eastdormpool-config)#class CLASS1
SwitchB(dhcpv6-pool-eastdormpool-class-class1-config)#address range 2001:da8:100:1::3
2001:da8:100:1::30
SwitchB(dhcpv6-pool-eastdormpool-class-class1-config)#exit
SwitchB(dhcpv6-eastdormpool-config)#class CLASS2
SwitchB(dhcpv6-pool-eastdormpool-class-class2-config)#address range 2001:da8:100:1::31
2001:da8:100:1::60
SwitchB(dhcpv6-eastdormpool-config)#class CLASS3
SwitchB(dhcpv6-pool-eastdormpool-class-class3-config)#address range 2001:da8:100:1::61
2001:da8:100:1::100
SwitchB(dhcpv6-pool-eastdormpool-class-class3-config)#exit
SwitchB(dhcpv6-eastdormpool-config)#exit
SwitchB(config)#interface vlan 1
SwitchB(config-if-vlan1)#ipv6 address 2001:da8:100:1::2/64
SwitchB(config-if-vlan1)#ipv6 dhcp server EastDormPool
SwitchB(config-if-vlan1)#exit
SwitchB(config)#
```

14.2.3.2 DHCPv6 Relay option37, 38 Example

Example 1：

When deploying IPv6 campus network, DHCPv6 server function of routing device can be

used for IPv6 address allocation if special server is used for uniform allocation and management for IPv6 address. DHCPv6 server supports both stateful and stateless DHCPv6.

Network topology:

In access layer, layer2 access device Switch1 connects users in dormitory; in first-level aggregation layer, aggregation device Switch2 is used as DHCPv6 relay agent; in second-level aggregation layer, aggregation device Switch3 is used as DHCPv6 server and connects with backbone network or devices in higher aggregation layer; in user side, PCs are generally loaded with Windows Vista system, thus having DHCPv6 client.

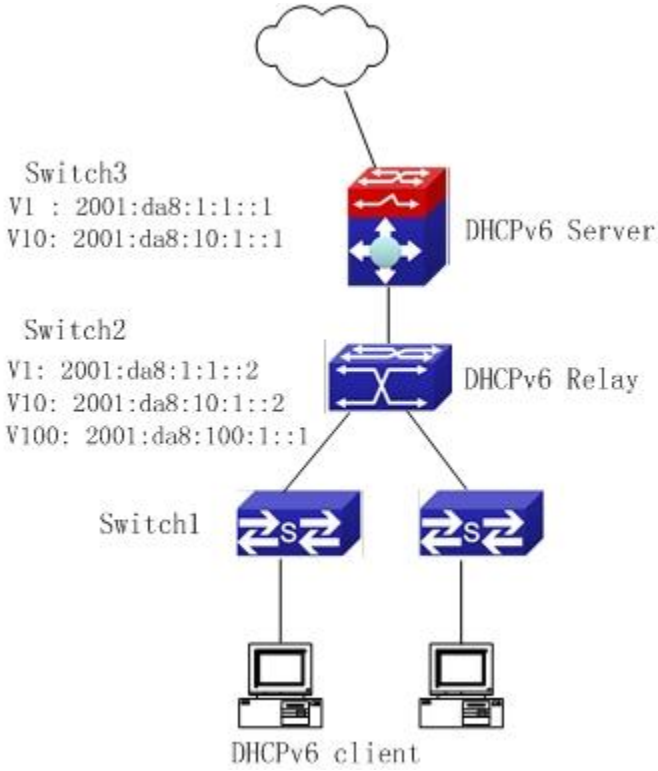


Figure 14-5 DHCPv6 relay option schematic

Switch2 configuration:

S2(config)#service dhcpv6

S2(config)#ipv6 dhcp relay remote-id option

S2(config)#ipv6 dhcp relay subscriber-id option

S2(config)#vlan 10

S2(config-vlan10)#int vlan 10

S2(config-if-vlan10)#ipv6 address 2001:da8:1:::2/64

S2(config-if-vlan10)#ipv6 dhcp relay destination 2001:da8:10:1::1

S2(config-if-vlan10)#exit

S2(config)#

14.2.4 DHCPv6 option37, 38 Troubleshooting

- Request packets sent by DHCPv6 client are multicast packets received by the device within its VLAN, if DHCPv6 server wants to receive the packets from client, DHCPv6 client and DHCPv6

server must be in the same VLAN, otherwise it needs to use DHCPv6 relay.

- ☞ Snooping option37,38 can process one of the following operations for DHCPv6 request packets with option37,38: replace the original option37,38 with its own; discard the packets with option37,38; do not execute adding, discarding or forwarding operation. Therefore, please check policy configuration of snooping option37,38 on second device when obtaining the false address or no address is obtained according to option37,38.
- ☞ DHCPv6 server obtains option37,38 of the packets from client by default, if no, it will obtain option37,38 of the packet sent by relay.
- ☞ DHCPv6 server only checks whether the first DHCPv6 relay adds option37,38 that means only option37,38 of the innermost relay-forw is valid in relay packets.

14.3 Prevent ND Spoofing

14.3.1 Overview

ND is neighbor discovering protocol in IPv6 protocol, and it’s similar to ARP on operation principle, therefore we do in the same way as preventing ARP spoofing to prevent ND spoofing and attack.

14.3.2 Prevent ND Spoofing configuration

The steps of preventingND spoofing configuration as below:

4. Disable ND automatic update function
5. Disable ND automatic learning function
6. Changing dynamic ND to static ND

1. Disable ND automatic update function

Command	Explanation
Global Mode and Port Mode	
ipv6 nd-security updateprotect **no ipv6 nd-security updateprotect**	Disable and enable ND automatic update function.

2. Disable ND automatic learning function

Command	Explanation
Global mode and Interface Mode	
ipv6 nd-security learnprotect **no ipv6 nd-security learnprotect**	Disable and enable ND automatic learning function.

3. Function on changing dynamic ND to static ND

Command	Explanation
Global Mode and Port Mode	
ipv6 nd-security convert	Change dynamic ND to static ND.

14.3.3 Prevent ND Spoofing Example

Please refer to the Arp binding example

14.4 RIPng

14.4.1 Introduction to RIPng

RIPng is first introduced in ARPANET, this is a protocol dedicated to small, simple networks. RIPng is a distance vector routing protocol based on the Bellman-Ford algorithm. Network devices running vector routing protocol send 2 kind of information to the neighboring devices regularly:

• Number of hops to reach the destination network, or metrics to use or number of networks to pass.

• What is the next hop, or the director (vector) to use to reach the destination network.

Distance vector layer3 switches send all their route selecting tables to the neighbor layer3 switches at regular interval. A layer3 switch will build their own route selecting information table based on the information received from the neighbor layer3 switches. Then, it will send this information to its own neighbor layer3 switches. As a result, the route selection table is built on second hand information, route beyond 15 hops will be deemed as unreachable.

RIPng is an optional routing protocol based on UDP. Hosts using RIPng send and receive packets on UDP port 521. All layer3 switches running RIP send their route table to all neighbor layer3 switches every 30 seconds for update. If no information from the partner is received in 180 seconds, then the device is deemed to have failed and the network connected to that device is considered to be unreachable. However, the route of that layer3 switch will be kept in the route table for another 120 seconds before deletion.

As layer3 switches running RIPng build route table with second hand information, infinite count may occur. For a network running RIPng routing protocol, when a RIPng route becomes unreachable, the neighboring RIPng layer3 switch will not send route update packets at once, instead, it waits until the update interval timeout (every 30 seconds) and sends the update packets containing that route. If before it receives the updated packet, its neighbors send packets

containing the information about the failed neighbor, “infinite count” will be resulted. In other words, the route of unreachable layer3 switch will be selected with the metrics increasing progressively. This greatly affects the route selection and route aggregation time.

To avoid “infinite count”, RIPng provides mechanism such as “split horizon” and “triggered update” to solve route loop. “Split horizon” is done by avoiding sending to a gateway routes leaned from that gateway. There are two split horizon methods: “simple split horizon” and “poison reverse split horizon”. Simple split horizon deletes from the route to be sent to the neighbor gateways the routes learnt from the neighbor gateways; poison reverse split horizon not only deletes the abovementioned routes, but set the costs of those routes to infinite. “Triggering update” mechanism defines whenever route metric changed by the gateway, the gateway advertise the update packets immediately other than wait for the 30 sec timer.

So far the RIPng protocol has got only one version----Version1: RIPng protocol is introduced in RFC 2080. RIPng transmits updating data packet by multicast data packet (multicast address FF02::9)

Each layer3 switch running RIPng has a route database, which contains all route entries for reachable destination, and route table is built based on this database. When a RIPng layer3 switch sent route update packets to its neighbor devices, the complete route table is included in the packets. Therefore, in a large network, routing data to be transferred and processed for each layer3 switch is quite large, causing degraded network performance.

Besides the above mentioned, RIPng protocol allows IPv6 route information discovered by the other routing protocols to be introduced to the route table.

The operation of RIPng protocol is shown below:

1. Enable RIPng The switch sends request packets to the neighbor layer3 switches by broadcasting; on receiving the request, the neighbor devices reply with the packets containing their local routing information.
2. The Layer3 switch modifies its local route table on receiving the reply packets and sends triggered update packets to the neighbor devices to advertise route update information. On receiving the triggered update packet, the neighbor lay3 switches send triggered update packets to their neighbor lay3 switches. After a sequence of triggered update packet broadcast, all layer3 switches get and maintain the latest route information.

In addition, RIPng layer3 switches will advertise its local route table to their neighbor devices every 30 seconds. On receiving the packets, neighbor devices maintain their local route table, select the best route and advertise the updated information to their own neighbor devices, so that the updated routes are globally valid. Moreover, RIP uses a timeout mechanism for outdated route, that is, if a switch does not receive regular update packets from a neighbor within a certain interval (invalid timer interval), it considers the route from that neighbor invalid, after holding the route fro a certain interval (garbage collect timer interval), it will delete that route.

As a result of continuous development of IPv6 network, it has the network environment of nonsupport IPv6 sometimes, so it needs to do the IPv6 operation by tunnel. Therefore, our RIPng supports configuration on configure tunnel, and passes through nonsupport IPv6 network by unicast packet of IPv4 encapsulation.

14.4.2 RIPng Configuration Task List

RIPng Configuration Task List:

1. Enable RIPng protocol (required)
   （1） Enable/disable RIPng protocol
   （2） Configure the interfaces running RIPng protocol
2. Configure RIPng protocol parameters (optional)
   （1） Configure RIPng sending mechanism
      1) Configure specified RIPng packets transmission address
   （2） Configure RIP routing parameters
      1) Configure route introduction (default route metric, configure routes of the other protocols to be introduced in RIPng)
      2) Configure the route deviation
      3) Configure and apply route filter
      4) Configure split horizon
3. Configure other RIPng parameters
   (2) Configure timer for RIPng update, timeout and hold-down
4. Delete the specified route in RIPng route table
5. Configure RIPng route aggregation
   （1） Configure aggregation route of IPv6 route mode
   （2） Configure aggregation route of IPv6 interface configuration mode
   （3） Display IPv6 aggregation route information
6. Configure redistribution of OSPFv3 routing to RIPng
   （1） Enable redistribution of OSPFv3 routing to RIPng
   （2） Display and debug the information about configuration of redistribution of OSPFv3 routing to RIPng

1. Enable RIPng protocol

Applying RIPng route protocol with basic configuration in switch is simple. Normally you only have to open the RIPng switch and configure the segments running RIPng, namely send and receive the RIPng data packet by default RIPng configuration.

Command	Explanation
Global mode	
[no] router IPv6 rip	Enables the RIPng protocol; the **no router IPv6 rip** command shuts the RIPng protocol.
Interface configuration mode	
[no] IPv6 router rip	Configure the interface to run RIPng protocol; the **no IPv6 router rip** command set the interface not run RIPng protocol.

2. Configure RIPng protocol parameters

（1）Configure RIPng sending mechanism

1）Configure the RIPng data packets point-transmitting

Command	Explanation
Router configuration mode	
[no] neighbor *<IPv6-address> <ifname>*	Specify the IPv6 Link-local address and interface of the neighboring route needs point-transmitting; the **no neighbor *<IPv6-address> <ifname>*** command cancels the appointed router.
[no] passive-interface *<ifname>*	Block the RIPng multicast on specified port and the RIPng data packet is only transmittable among Layer 3 switch configured with neighbor. The **no passive-interface *<ifname>*** command cancels the function.

（2）Configure RIP routing parameters

1）Configure route introduction (default route metric, configure routes of the other protocols to be introduced in RIP)

Command	Explanation
Router configuration mode	
default-metric *<value>* **no default-metric**	Configure the default metric of distributed route; the **no default-metric** command restores the default configuration 1.
[no]redistribute {kernel |connected| static| ospf| isis| bgp} [metric<*value*>] [route-map<*word*>]	Redistribute the routes distributed in other route protocols into the RIPng data packet; the **no redistribute {kernel |connected| static| ospf| isis| bgp} [metric<*value*>] [route-map<*word*>]** command cancels the distributed route of corresponding protocols.
[no]default-information originate	Generate a default route to the RIPng protocol; the **no default-information originate** command cancels the feature.

2）Configure the route offset

Command	Explanation
Router configuration mode	
[no] offset-list <access-list-number |access-list-name> {in|out} <number > [<ifname>]	Configure that provide a deviation value to the route metric value when the port sends or receives RIPng data packet; the **no offset-list *<access-list-number |access-list-name>* {in|out} *<number >* [*<ifname>*]** command removes the deviation table.

3）Configure and apply route filter and route aggregation

Command	Explanation
Router configuration mode	

[no] distribute-list {*<access-list-number /access-list-name>* | prefix*<prefix-list-name>*} {in|out} [*<ifname>*]	Set to filter the route when the interface sends and receives RIPng data packets. The **no distribute-list {*< access-list-number /access-list-name >* | prefix*<prefix-list-name>*} {in|out} [*<ifname>*]** command means do not set the route filter.
[no]aggregate-address *<IPv6-address>*	Configure route aggregation, the **no aggregate-address *<IPv6-address*** command cancels the route aggregation.

4）Configure split horizon

Command	Explanation
Interface configuration mode	
IPv6 rip split-horizon [poisoned]	Configure that take the split-horizon when the port sends data packets, **poisoned** means with poison reverse.
no IPv6 rip split-horizon	Cancel the split-horizon.

3. Configure other RIPng protocol parameters

(1) Configure timer for RIPng update, timeout and hold-down

Command	Explanation
Router configuration mode	
timers basic *<update> <invalid> <garbage>* **no timers basic**	Adjust update, timeout and garbage recycle of RIPng timer, the **no timers basic** command restores the default configuration.

4. Delete the specified route in RIPng route table

Command	Explanation
Admin Mode	
clear IPv6 rip route {*<IPv6-address>*|kernel|static|connected|rip|ospf|isis|bgp|all}	The command deletes a specified route from the RIP route table.

5. Configure RIPng route aggregation

(1) Configure IPv6 aggregation route globally

Command	Explanation
Router Configuration Mode	
ipv6 rip aggregate-address X:X::X:X/M **no ipv6 rip aggregate-address X:X::X:X/M**	To configure or delete IPv6 aggregation route globally.

(2) Configure IPv6 aggregation route on interface

Command	Explanation

Interface Configuration Mode	
ipv6 rip aggregate-address X:X::X:X/M **no ipv6 rip aggregate-address X:X::X:X/M**	To configure or delete IPv6 aggregation route on interface.

(3) Display IPv6 aggregation route information

Command	Explanation
Admin Mode and Configuration Mode	
show ipv6 rip aggregate	To display IPv6 aggregation route information, such as aggregation interface, metric, numbers of aggregation route, times of aggregation.

6. Configure redistribution of OSPFv3 routing to RIPng

(1)Enable redistribution of OSPFv3 routing to RIPng

Command	Explanation
Router IPv6 RIP Configuration Mode	
redistribute ospf [<*process-tag*>] [metric<*value*>] [route-map<*word*>] **no redistribute ospf [<*process-tag*>]**	To enable or disable redistribution of OSPFv3 routing for RIPng.

(2) Display and debug the information about configuration of redistribution of OSPFv3 routing to RIPng

Command	Explanation
Admin Configuration Mode	
show ipv6 rip redistribute	To display RIPng routing which is redistributed from other routing protocols.
Admin Mode	
debug ipv6 rip redistribute message send **no debug ipv6 rip redistribute message send** **debug ipv6 rip redistribute route receive** **no debug ipv6 rip redistribute route receive**	To enable or disable debugging messages sent by RIPng for redistribution of OSPFv3 routing. To enable or disable debugging route messages received from NSM.

14.4.3 RIPng Configuration Examples

14.4.3.1 Typical RIPng Examples

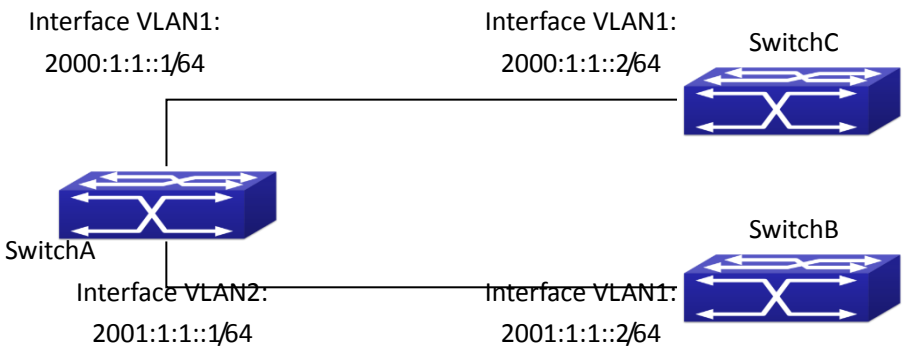


Figure 14-6 RIPng Example

As shown in the above figure, a network consists of three layer 3 switches. SwitchA and SwitchB connect to SwitchC through interface vlan1 and vlan2. All the three switches are running RIPng. Assume SwitchA (VLAN1：2001:1:1::1/64 and VLAN2：2001:1:1::1/64) exchange update information with SwitchB (VLAN1：2001:1:1::2/64) only, update information is not exchanged between SwitchA and SwitchC (VLAN1：2001:1:1::2/64).

The configuration for SwitchA, SwitchB and SwitchC is shown below:

Layer 3 SwitchA

Enable RIPng protocol

SwitchA(config)#router IPv6 rip

SwitchA(config-router)#exit

Configure the IPv6 address in vlan1 and configure vlan1 to run RIPng

SwitchA#config

SwitchA(config)# interface Vlan1

SwitchA(config-if-Vlan1)# IPv6 address 2000:1:1::1/64

SwitchA(config-if-Vlan1)#IPv6 router rip

SwitchA(config-if-Vlan1)#exit

Configure the IPv6 address in vlan2 and configure vlan2 to run RIPng

SwitchA(config)# interface Vlan2

SwitchA(config-if-Vlan2)#IPv6 address 2001:1:1::1/64

SwitchA(config-if-Vlan2)#IPv6 router rip

SwitchA(config-if-Vlan2)#exit

Configure the interface vlan1 do not send RIPng messages to SwitchC

SwitchA(config)#

SwitchA(config-router)#passive-interface Vlan1

SwitchA(config-router)#exit

Layer 3 SwitchB

Enable RIPng protocol

SwitchB (config)#router IPv6 rip

SwitchB (config-router-rip)#exit

Configure the IPv6 address and interfaces of Ethernet port vlan1 to run RIPng

SwitchB#config

SwitchB(config)# interface Vlan1

SwitchB(config-if)# IPv6 address 2001:1:1::2/64
SwitchB(config-if)#IPv6 router rip
SwitchB(config-if)exit

Layer 3 SwitchC
Enable RIPng protocol
SwitchC(config)#router IPv6 rip
SwitchC(config-router-rip)#exit
Configure the IPv6 address and interfaces of Ethernet port vlan1 to run RIPng
SwitchC#config
SwitchC(config)# interface Vlan1
SwitchC(config-if)# IPv6 address 2000:1:1::2/64
SwitchC(config-if)#IPv6 router rip
SwitchC(config-if)exit

14.4.3.2 RIPng Aggregation Route Function Typical Examples

The application topology as follows:

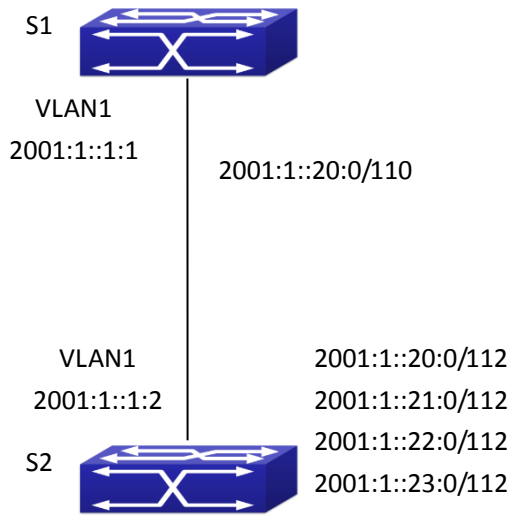


Figure 14-7 Typical application of RIPng aggregation

As the above network topology, S2 is connected to S1 through interface vlan1, there are other 4 subnet routers of S2, which are 2001:1::20:0/112, 2001:1::21:0/112, 2001:1::22:0/112, 2001:1::23:0/112. S2 supports route aggregation, and to configure aggregation route 2001:1::20:0/110 in interface vlan1 of S2, after that, sending router messages to S2 through vlan1, and put the four subnet routers aggregated to one router as 2001:1::20:0/110, and send to S1, and not send subnet to neighbor. It can reduce the router table of S1, save the memory.

S1 configuration list:
S1(config)#router ipv6 rip
S1(config)# interface Vlan1
S1(config-if-Vlan1)#IPv6 address 2001:1::1:1/112
S1(config-if-Vlan1)#IPv6 router rip

S2 configuration list:
S2(config)#router ipv6 rip
S2(config)#interface vlan 1
S2(config-if-Vlan1)#IPv6 address 2001:1::1:2/112
S2(config-if-Vlan1)#IPv6 router rip
S2(Config-if-Vlan1)#ipv6 rip agg 2001:1::20:0/110

14.4.4 RIPng Troubleshooting

The RIPng protocol may not be working properly due to errors such as physic connection, configuration error when configuring and using the RIPng protocol. So users should pay attention to the following:

- First ensure the physic connection is correct and the IP Forwarding command is open
- Second, ensure the interface and link layer protocol are UP (use **show interface** command)
- Then initiate the RIPng protocol (use **router IPv6 rip** command) and configure the port (use **IPv6 router** command), and set RIPng protocol parameter on corresponding interfaces.
- After that, a RIPng protocol feature should be noticed ---the Layer 3 switch running RIPng transmits the route updating messages every 30 seconds. A Layer 3 switch is considered inaccessible if no route updating messages from the switch are received within 180 seconds, then the route to the switch will remains in the route table for 120 seconds before it is deleted. Therefore, if to delete a RIPng route, this route item is assured to be deleted from route table after 300 seconds.
- If the RIP routing problem remains unresolved, please use **debug IPv6 rip** command to record the debug message in three minutes, and send them to our technical service center.

14.5 OSPFv3

14.5.1 Introduction to OSPFv3

OSPFv3 (Open Shortest Path First) is the third version for Open Shortest Path First, and it is the IPv6 version of OSPF Protocol. It is an interior dynamic routing protocol for autonomous system based on link-state. The protocol creates a link-state database by exchanging link-states among layer3 switches, and then uses the Shortest Path First algorithm to generate a route table basing on that database.

Autonomous system (AS) is a self-managed interconnected network. In large networks, such as the Internet, a giant interconnected network is broken down to autonomous systems. Big enterprise networks connecting to the Internet are independent AS, since the other hosts on the Internet are not managed by those AS and they don’t share interior routing information with the

layer3 switches on the Internet.

Each link-state layer3 switch can provide information about the topology with its neighboring layer3 switches.

• The network segment (link) connecting to the layer3 switch

• State of the connecting link

Link-state information is flooded throughout the network so that all layer3 switches can get first hand information. Link-state layer3 switches will not broadcast all information contained in their route tables; instead, they only send changed link-state information. Link-state layer3 switches establish neighborhood by sending “HELLO” to their neighbors, then link-state advertisements (LSA) will be sent among neighboring layer3 switches. Neighboring layer3 switch copy the LSA to their routing table and transfer the information to the rest part of the network. This process is referred to as “flooding”. In this way, firsthand information is sent throughout the network to provide accurate map for creating and updating routes in the network. Link-state routing protocols use cost instead of hops to decide the route. Cost is assigned automatically or manually. According to the algorithm in link-state protocol, cost can be used to calculate the hop number for packets to pass, link bandwidth, and current load of the link, the administrator can even add weight for better assessment of the link-state.

1) When a link-state layer3 switch enters a link-state interconnected network, it sends a HELLO packet to get to know its neighbors and establish neighborhood.

2) The neighbors respond with information about the links they are connecting and the related costs.

3) The originate layer3 switch uses this information to build its own routing table.

4) Then, as part of the regular update, layer3 switch send link-state advertisement (LSA) packets to its neighboring layer3 switches. The LSA include links and related costs of that layer3 switch.

5) Each neighboring layer3 switch copies the LSA packet and passes it to the next neighbor (i.e. flooding).

6) Since routing database is not recalculated before layer3 switch forwards LSA flooding, the converging time is greatly reduced.

One major advantage of link-state routing protocols is the fact that infinite counting is impossible, this is because of the way link-state routing protocols build up their routing table. The second advantage is that converging in a link-state interconnected network is very fast, once the routing topology changes, updates will be flooded throughout the network very soon. Those advantages release some layer3 switch resources, as the process ability and bandwidth used by bad route information are minor.

The features of OSPFv3 protocol include the following: OSPFv3 supports networks of various scales, several hundreds of layer3 switches can be supported in an OSPFv3 network. Routing topology changes can be quickly found and updating LSAs can be sent immediately, so that routes converge quickly. Link-state information is used in shortest path algorithm for route calculation, eliminating loop route. OSPFv3 divides the autonomous system intro areas, reducing database size, bandwidth occupation and calculation load. (According to the position of layer3 switches in the autonomous system, they can be grouped as internal area switches, area edge switches, AS edge switches and backbone switches). OSPFv3 supports load balance and multiple routes to the same destination of equal costs. OSPFv3 supports 4 level routing mechanisms (process routing

according to the order of route inside an area, route between areas, type 1 external route and type 2 external route). OSPFv3 support IP subnet and redistribution of routes from the other routing protocols, and interface-based packet verification. OSPFv3 supports sending packets in multicast.

Each OSPFV3 layer3 switch maintains a database describing the topology of the whole autonomous system. Each layer3 switch gathers the local status information, such as available interface, reachable neighbors, and sends link-state advertisement (sending out link-state information) to exchange link-state information with other OSPFv3 layer3 switches to form a link-state database describing the whole autonomous system. Each layer3 switch builds a shortest path tree rooted by itself according to the link-state database, this tree provide the routes to all nodes in an autonomous system. If two or more layer3 switches exist (i.e. multi-access network), "designated layer3 switch" and "backup designated layer3 switch" will be selected. Designated layer3 switch is responsible for spreading link-state of the network. This concept helps reducing the traffic among the Layer3 switches in multi-access network.

OSPFv3 protocol requires the autonomous system to be divided into areas. That is to divide the autonomous system into 0 area (backbone area) and non-0 areas. Routing information between areas are further abstracted and summarized to reduce the bandwidth required in the network. OSPFv3 uses four different kinds of routes: they are the route inside the area, route between areas, type 1 external route and type 2 external route, in the order of highest priority to lowest. The route inside an area and between areas describe the internal network structure of an autonomous system, while external routes describe external routes describe how to select the routing information to destination outside the autonomous system. The first type of exterior route corresponds to the information introduced by OSPFv3 from the other interior routing protocols, the costs of those routes are comparable with the costs of OSPFv3 routes; the second type of exterior route corresponds to the information introduced by OSPFv3 from the other exterior routing protocols, but the costs of those routes are far greater than that of OSPFv3 routes, so OSPFv3 route cost is ignored when calculating route costs.

OSPFv3 areas are centered with the Backbone area, identified as the Area 0, all the other areas must be connected to Area 0 logically, and Area 0 must be continuous. For this reason, the concept of virtual link is introduced to the backbone area, so that physically separated areas still have logical connectivity to the backbone area. The configurations of all the layer3 switches in the same area must be the same.

In one word, LSA can only be transferred between neighboring Layer3 switches, and OSPFv3 protocol includes seven kinds of LSA: link LSA, internal-area prefix LSA, router LSA, network LSA, inter-area prefix LSA, inter-area router LSA and autonomic system exterior LSA. Router LSA is generated by each Layer 3 switch in an OSPF area, and is sent to all other neighboring Layer 3 switch in this area; network LSA is generated by designated Layer 3 switch in the OSPF area of multi-access network and is sent to all other neighboring layer3 switches in this area.(To reduce data traffic among each Layer 3 switches in the multi-access network, "designated layer3 switch" and "backup designated layer3 switch" should be selected in the multi-access network, and the network link-state is broadcasted by designated Layer 3 switch); the inter-area prefix LSA and inter-area router LSA are generated by OSPF area border Layer 3 switches and transferred among those switches. The autonomic system exterior LSA is generated by autonomic system exterior border Layer 3 switches and transferred in the whole autonomic

system. Link LSA is generated by Layer 3 switch on the link and sent to other Layer 3 switches on the link. Internal-area prefix LSA is generated by designated layer3 switch of each link in this area, and flooded to the whole area.

For autonomous system focused on exterior link-state announcement, OSPFv3 allow some areas to be configured as STUB areas in order to reduce the size of topological database. Router LSA, network LSA, inter-area prefix LSA, link LSA, internal-area prefix LSA are permitted to advertise to STUB area. Default route must be used in STUB area, Layer 3 switches on the area border of STUB area announces to default routes of STUB area by inter-area prefix LSA; these default routes only flood in STUB area, not outside of STUB area. Each STUB area has a corresponding default route, the route from STUB area to AS exterior destination depends only on default route of this area.

The following simply outlines the route calculation process of OSPFv3 protocol:

1) Each OSPF-enabled layer3 switch maintains a database (LS database) describing the link-state of the topology structure of the whole autonomous system. Each layer3 switch generates a link-state advertisement according to its surrounding network topology structure (router LSA), and sends the LSA to other layer3 switches through link-state update (LSU) packets. Thus, each layer3 switches receives LSAs from other layer3 switches, and all LSAs combined to the link-state database.
2) Since a LSA is the description of the network topology structure around a layer3 switch, the LS database is the description of the network topology structure of the whole network. The layer3 switches can easily create a weighted vector map according to the LS database. Obviously, all layer3 switches in the same autonomous system will have the same network topology map.
3) Each layer3 switch uses the shortest path first (SPF) algorithm to calculate a tree of shortest path rooted by itself. The tree provides the route to all the nodes in the autonomous system, leaf nodes consist of the exterior route information. The exterior route can be marked by the layer3 switch broadcast it, so that additional information about the autonomous system can be recorded. As a result, the route table of each layer3 switch is different.

OSPFv3 protocol is developed by the IETF, the OSPF v3 used now is fulfilled according to the content described in RFC2328 and RFC2740.

As a result of continuous development of IPv6 network, it has the network environment of nonsupport IPv6 sometimes, so it needs to do the IPv6 operation by tunnel. Therefore, our OSPFv3 supports configuration on configure tunnel, and passes through nonsupport IPv6 network by unicast packet of IPv4 encapsulation.

14.5.2 OSPFv3 Configuration Task List

OSPFv3 Configuration Task List:

1. Enable OSPFv3 (required)
   （1） Enable/disable OSPFv3 (required)
   （2） Configure the router-id number of the layer3 switch running OSPFv3 (optional)

（3） Configure the network scope for running OSPFv3 (optional)

（4） Enable OSPFv3 on the interface (required)

2. Configure OSPFv3 auxiliary parameters (optional)

（1） Configure OSPFv3 packet sending mechanism parameters

1) Set the OSPFv3 interface to receive only

2) Configure the cost for sending packets from the interface

3) Configure OSPFv3 packet sending timer parameter (timer of broadcast interface sending HELLO packet to poll, timer of neighboring layer3 switch invalid timeout, timer of LSA transmission delay and timer of LSA retransmission).

（2） Configure OSPFv3 route introduction parameters

1) Configure default parameters (default type, default tag value, default cost)

2) Configure the routes of the other protocols to introduce to OSPFv3

（3） Configure OSPFv3 importing the routes of other OSPFv3 processes

1) Enable the function of OSPFv3 importing the routes of other OSPFv3 processes

2) Display relative information

3) Debug

（4） Configure other OSPFv3 protocol parameters

1) Configure OSPFv3 routing protocol priority

2) Configure cost for OSPFv3 STUB area and default route

3) Configure OSPFv3 virtual link

4) Configure the priority of the interface when electing designated layer3 switch

3. Close OSPFv3 Protocol

1. Enable OSPFv3 Protocol

It is very simple to run the basic configurations of OSPFv3 routing protocol on the Layer 3 switch, normally only enabling OSPFv3, implement OSPFv3 interface, the default value is defined to OSPFv3 protocol parameters. Refer to 2. Configure OSPF auxiliary parameters, if the OSPFv3 protocol parameters need to be modified.

Commands	Explanation
Global Mode	
[no] router IPv6 ospf <*tag*>	The command initializes OSPFv3 routing process and enter OSPFv3 mode to configure OSPFv3 routing process. The **no router IPv6 ospf <*tag*>** command stops relative process. (required)
OSPFv3 Protocol Configure Mode	
router-id <*router_id*> **no router-id**	Configure router for OSPFv3 process. The **no router-id** command returns ID to 0.0.0.0 .(required)

[no] passive-interface<*ifname*>	Configure an interface receiving without sending. The **no passive-interface<*ifname*>** command cancels configuration.
Interface Configuration Mode	
[no] IPv6 router ospf {area <*area-id*> [instance-id <*instance-id*> | tag <*tag*> [instance-id <*instance-id*>]] | tag <*tag*> area <*area-id*> [instance-id <*instance-id*>]}	Implement OSPFv3 routing on the interface. The **no IPv6 router ospf {area <*area-id*> [instance-id <*instance-id*> | tag <*tag*> [instance-id <*instance-id*>]] | tag <*tag*> area <*area-id*> [instance-id <*instance-id*>]}** command cancels configuration.

2. Configure OSPFv3 parameters

（1）Configure OSPFv3 packet sending mechanism parameters

1）Set the OSPF interface to receive only

2）Configure the cost for sending packets from the interface

Commands	Explanation
Interface Configuration Mode	
IPv6 ospf cost <*cost*> [instance-id <*id*>] **no IPv6 ospf cost [instance-id <*id*>]**	Appoint interface to implement required cost of OSPFv3 protocol. The **no IPv6 OSPF cost [instance-id <*id*>]** restores the default setting.

3）Configure OSPFv3 packet sending timer parameter (timer of broadcast interface sending HELLO packet to poll, timer of neighboring layer3 switch invalid timeout, timer of LSA transmission delay and timer of LSA retransmission).

Commands	Explanation
Interface Configuration Mode	
IPv6 ospf hello-interval <*time*> [instance-id <*id*>] **no IPv6 ospf hello-interval [instance-id <*id*>]**	Sets interval for sending HELLO packets; the **no IPv6 ospf hello-interval [instance-id <*id*>]** command restores the default setting.
IPv6 ospf dead-interval <*time*> [instance-id <*id*>] **no IPv6 ospf dead-interval [instance-id <*id*>]**	Sets the interval before regarding a neighbor layer3 switch invalid; the **no IPv6 ospf dead-interval [instance-id <*id*>]** command restores the default setting.
IPv6 ospf transit-delay <*time*> [instance-id <*id*>] **no IPv6 ospf transit-delay [instance-id <*id*>]**	Sets the delay time before sending link-state broadcast; the **no IPv6 ospf transit-delay [instance-id <*id*>]** command restores the default setting.
IPv6 ospf retransmit <*time*> [instance-id <*id*>] **no IPv6 ospf retransmit [instance-id <*id*>]**	.Sets the interval for retransmission of link-state advertisement among neighbor layer3 switches; the **no IPv6 ospf retransmit [instance-id <*id*>]** command restores the default setting.

（2）Configure OSPFv3 route introduction parameters

Configure OSPFv3 route introduction parameters

Commands	Explanation
OSPF Protocol Mode	
[no]redistribute {kernel |connected| static| rip| isis| bgp} [metric<*value*>] [metric-type {1|2}][route-map<*word*>]	Introduces other protocol discovery routing and static routing regarded as external routing message. The **no redistribute {kernel |connected| static| rip| isis| bgp} [metric<*value*>] [metric-type {1|2}][route-map<*word*>]** command cancels imported external routing message.

（3）Configure OSPFv3 importing the routes of other OSPFv3 processes

1）Enable the function of OSPFv3 importing the routes of other OSPFv3 processes

Command	Explanation
Router IPv6 OSPF Mode	
redistribute ospf [<*process-id*>] [metric<*value*>] [metric-type {1|2}][route-map<*word*>] **no redistribute ospf [<*process-id*>] [metric<*value*>] [metric-type {1|2}][route-map<*word*>]**	Enable or disable the function of OSPFv3 importing the routes of other OSPFv3 processes.

2）Display relative information

Command	Explanation
Admin Mode or Configure Mode	
show ipv6 ospf [<*process-id*>] redistribute	Display the configuration information of the OSPFv3 process importing other outside routes.

3）Debug

Command	Explanation
Admin Mode	
debug ipv6 ospf redistribute message send **no debug ipv6 ospf redistribute message send** **debug ipv6 ospf redistribute route receive** **no debug ipv6 ospf redistribute route receive**	Enable or disable debugging of sending command from OSPFv3 process redistributed to other OSPFv3 process routing. Enable or disable debugging of received routing message from NSM for OSPFv3 process.

（4） Configure Other Parameters of OSPFv3 Protocol

1） Configure OSPFv3 STUB Area & Default Routing Cost

2） Configure OSPFv3 Virtual Link

Commands	Explanation
OSPFv3 Protocol Configuration Mode	

timers spf *<spf-delay>* *<spf-holdtime>* **no timers spf**	Configure OSPFv3 SPF timer. The **no timers spf** command recovers default value.
area <*id*> stub [no-summary] **no area <*id*> stub [no-summary]** **area <*id*> default-cost <*cost*>** **no area <*id*> default-cost** **area <*id*> virtual-link A.B.C.D [instance-id <*instance-id*> INTERVAL]** **no area <*id*> virtual-link A.B.C.D [|INTERVAL]**	Configure parameters in OSPFv3 area (STUB area, Virtual link). The no command restores default value.

4）Configure the priority of the interface when electing designated layer3 switch (DR).

Commands	Explanation
Interface Configuration Mode	
IPv6 ospf priority <*priority*> [instance-id <*id*>] **no IPv6 ospf priority [instance-id <*id*>]**	Sets the priority of the interface in "designated layer3 switch" election; the "**no IPv6 ospf priority [instance-id <*id*>]**" command restores the default setting.

3. Disable OSPFv3 Protocol

Commands	Explanation
Global Mode	
no router IPv6 ospf ospf [<*tag*>]	Disable OSPFv3 Routing Protocol.

14.5.3 OSPFv3 Examples

Examples 1: OSPF autonomous system.

This scenario takes an OSPF autonomous system consists of five switch for example.

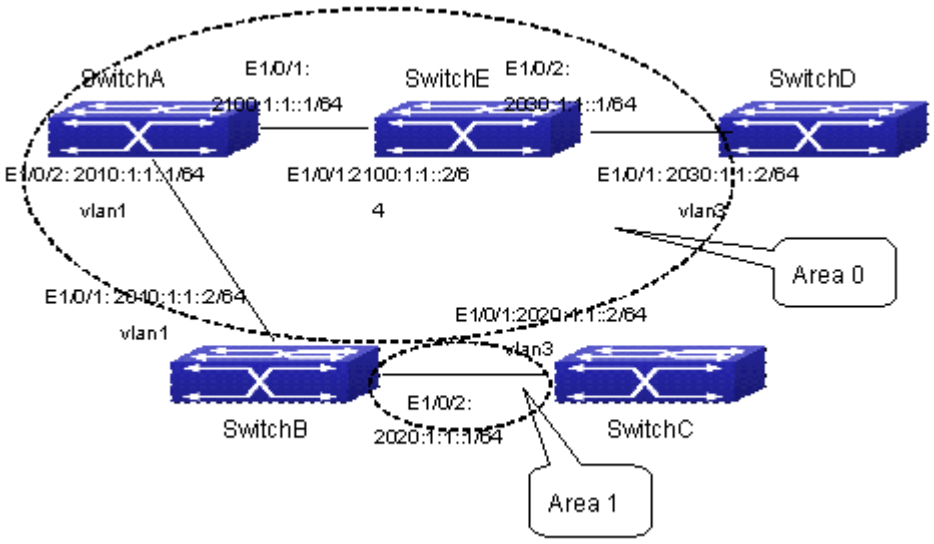


Figure 14-8 Network topology of OSPF autonomous system

The configuration for layer3 SwitchA and SwitchE is shown below:

Layer3 SwitchA:

Enable OSPFv3 protocol, configure router ID

SwitchA(config)#router IPv6 ospf

SwitchA (config-router)#router-id 192.168.2.1

Configure interface vlan1 IPv6 address and affiliated OSPFv3 area

SwitchA#config

SwitchA(config)# interface vlan 1

SwitchA(config-if-vlan1)# IPv6 address 2010:1:1::1/64

SwitchA(config-if-vlan1)# IPv6 router ospf area 0

SwitchA(config-if-vlan1)#exit

Configure interface vlan2 IP address and affiliated OSPFv3 area

SwitchA(config)# interface vlan 2

SwitchA(config-if-vlan2)# IPv6 address 2100:1:1::1/64

SwitchA(config-if-vlan2)# IPv6 router ospf area 0

SwitchA (config-if-vlan2)#exit

SwitchA(config)#exit

SwitchA#

Layer 3 SwitchB:

Enable OSPFv3 protocol, configure router ID

SwitchB(config)#router IPv6 ospf

SwitchB (config-router)#router-id 192.168.2.2

Configure interface vlan1 address, VLAN2 IPv6 address and affiliated OSPFv3 area

SwitchB#config

SwitchB(config)# interface vlan 1

SwitchB(config-if-vlan1)# IPv6 address 2010:1:1::2/64

SwitchB(config-if-vlan1)# IPv6 router ospf area 0

SwitchB(config-if-vlan1)#exit

SwitchB(config)# interface vlan 3

SwitchB(config-if-vlan3)# IPv6 address 2020:1:1::1/64
SwitchB(config-if-vlan3)# IPv6 router ospf area 1
SwitchB(config-if-vlan3)#exit
SwitchB(config)#exit
SwitchB#
Layer 3 SwitchC:
Enable OSPFv3 protocol, configure router ID
SwitchC(config)#router IPv6 ospf
SwitchC(config-router)#router-id 192.168.2.3
Configure interface vlan3 IPv6 address and affiliated OSPFv3 area
SwitchC#config
SwitchC(config)# interface vlan 3
SwitchC(config-if-vlan3)# IPv6 address 2020:1:1::2/64
SwitchC(config-if-vlan3)# IPv6 router ospf area 1
SwitchC(config-if-vlan3)#exit
SwitchC(config)#exit
SwitchC#
Layer 3 SwitchD:
Enable OSPFv3 protocol, configure router ID
SwitchD(config)#router IPv6 ospf
SwitchD(config-router)#router-id 192.168.2.4
Configure interface vlan3 IPv6 address and affiliated OSPFv3 area
SwitchD#config
SwitchD(config)# interface vlan 3
SwitchD(config-if-vlan3)# IPv6 address 2030:1:1::2/64
SwitchD(config-if-vlan3)# IPv6 router ospf area 0
SwitchD(config-if-vlan3)#exit
SwitchD(config)#exit
SwitchD#
Layer 3 SwitchE:
Startup OSPFv3 protocol, configure router ID
SwitchE(config)#router IPv6 ospf
SwitchE(config-router)#router-id 192.168.2.5
Configure interface IPv6 address and affiliated OSPFv3 area
SwitchE#config
SwitchE(config)# interface vlan 2
SwitchE(config-if-vlan2)# IPv6 address 2100:1:1::2/64
SwitchE(config-if-vlan2)# IPv6 router ospf area 0
SwitchE(config-if-vlan2)#exit
Configure interface VLAN3 IPv6 address and affiliated area
SwitchE(config)# interface vlan 3
SwitchE(config-if-vlan3)# IPv6 address 2030:1:1::1/64
SwitchE(config-if-vlan3)# IPv6 router ospf area 0
SwitchE(config-if-vlan3)#exit

SwitchE(config)#exit
SwitchE#

14.5.4 OSPFv3 Troubleshooting

In the process of configuring and implementing OSPFv3, physical connection, configuration false probably leads to OSPFv3 protocol doesn't work. Therefore, the customers should give their attention to it:

- First of all, to ensure correct physical connection;
- Secondly, to ensure interface and link protocol are UP (execute **show interface** instruction);
- And configure IPv6 address of the different net segment on every interface.
- To startup OSPFv3 protocol (execute **router IPv6 OSPF** instruction), and configure affiliated OSPFv3 area on relative interface.
- And then, consider OSPFv3 protocol characteristic —— OSPFv3 backbone area (area 0) must be continuous. If it doesn't ensure that virtual link is implemented continuously, all of not area 0 only can be connected by area 0 and other not area 0, not directly connected by not area 0; The border Layer 3 switch is a part of this Layer 3 switch interface belongs to area 0, and another part of interface belongs to not area 0; for multi-access net etc like broadcast, Layer 3 switch DR needs vote and appoint; for each OSPFv3 process must not configure router ID of 0.0.0.0 address.

14.6 BGP4+

14.6.1 Introduction to MBGP4+

MBGP4+ is multi-protocol BGP (Multi-protocol Border Gateway Protocol) extension to IPv6, referring to BGP protocol chapter about BGP protocol introduction in this manual. Different from RIPng and OSPFv3, BGP has no corresponging independent protocol for IPv6; instead, it takes extensions to address families on the original BGP. The extensions to BGP by MBGP4+ are mostly embodied:

a. neighbor address configured can be IPv6 address;

b. Increase IPv6 unicast address family configuration.

14.6.2 MBGP4+ Configuration Task List

MBGP4+ Configuration Task List:

1. Configure IPv6 neighbor

2. Configure and enable IPv6 address family
3. Configure redistribution of OSPFv3 routing to MBGP4+
   1) Enable redistribution of OSPFv3 routing to MBGP4+
   2) Display and debug the information about configuration of redistribution of OSPFv3 routing to MBGP4+

1. Configure IPv6 neighbor

Command	Explanation
BGP Protocol Configuration Mode	
neighbor *<X:X::X:X>* remote-as *<as-id>*	Configure IPv6 neighbor.

2. Configure and activate IPv6 address family

Command	Explanation
BGP Protocol Configuration Mode	
address-family IPv6 unicast	Enter IPv6 unicast address family.
BGP protocol address family comfiguration mode	
neighbor *<X:X::X:X>* activate **no neighbor *<X:X::X:X>* activate**	Configure IPv6 neighbor to activate/inactivate the address family.
exit-address-family	Exit address family configuration mode.

3. Configure redistribution of OSPFv3 routing to MBGP4+

(1) Enable redistribution of OSPFv3 routing to MBGP4+

Command	Explanation
Router IPv6 BGP Configuration Mode	
redistribute ospf [*<process-tag>*] [route-map*<word>*] **no redistribute ospf [*<process-tag>*]**	To enable or disable redistribution of OSPFv3 routing to MBGP4+.

(2) Display and debug the information about configuration of redistribution of OSPFv3 routing to MBGP4+

Command	Explanation
Admin Mode and Configuration Mode	
show ipv6 bgp redistribute	To display configuration information about MBGP4+ routing which is redistributed from other routing protocols.
Admin Mode	
debug ipv6 bgp redistribute message send **no debug ipv6 bgp redistribute message send** **debug ipv6 bgp redistribute route receive** **no debug ipv6 bgp redistribute route receive**	To enable or disable debugging messages sent by MBGP4+ for redistribution of OSPFv3 routing. To enable or disable debugging messages received from NSM.

14.6.3 MBGP4+ Examples

SwitchB, SwitchC and SwitchD are in AS200, SwitchA is in AS100. SwitchA and SwitchB share the same network segment. SwitchB and SwitchD are not connected physically.

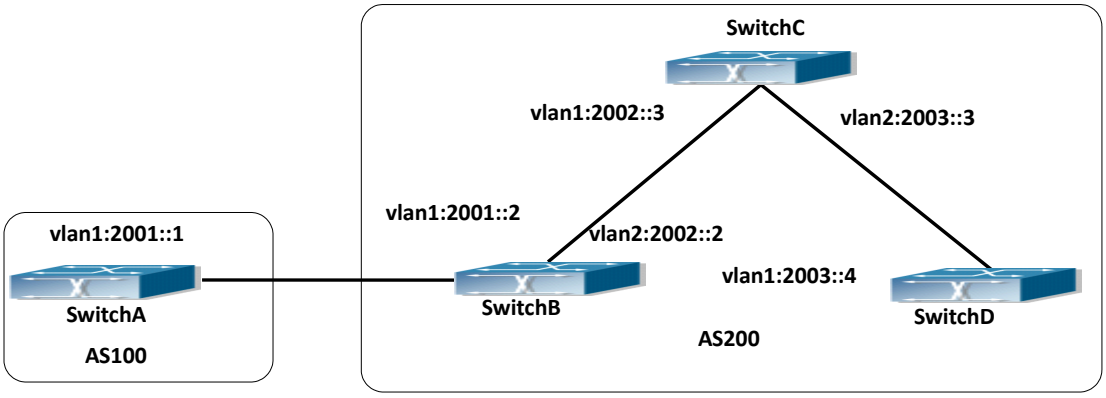


Figure 14-9 BGP Network Topological Map

Accordingly SwitchA configuration as follows:

SwitchA(config)#router bgp 100
SwitchA(config-router)#bgp router-id 1.1.1.1
SwitchA(config-router)#neighbor 2001::2 remote-as 200
SwitchA(config-router)#address-family IPv6 unicast
SwitchA(config-router-af)#neighbor 2001::2 activate
SwitchA(config-router-af)#exit-address-family
SwitchA(config-router-bgp)#exit
SwitchA(config)#

SwitchB configuration as follows:

SwitchB(config)#router bgp 200
SwitchA(config-router)#bgp router-id 2.2.2.2
SwitchB(config-router)#neighbor 2001::1 remote-as 100
SwitchB(config-router)#neighbor 2002::3 remote-as 200
SwitchB(config-router)#neighbor 2003::4 remote-as 200
SwitchB(config-router)#address-family IPv6 unicast
SwitchB(config-router-af)#neighbor 2001::1 activate
SwitchB(config-router-af)#neighbor 2002::3 activate
SwitchB(config-router-af)#neighbor 2003::4 activate
SwitchB(config-router-af)#exit-address-family
SwitchB(config-router)#exit
SwitchB(config)#

SwitchC configuration as follows:

SwitchC(config)#router bgp 200
SwitchA(config-router)#bgp router-id 2.2.2.2
SwitchC(config-router)#neighbor 2002::2 remote-as 200
SwitchC(config-router)#neighbor 2003::4 remote-as 200

SwitchC(config-router)#address-family IPv6 unicast
SwitchC(config-router-af)#neighbor 2002::2 activate
SwitchC(config-router-af)#neighbor 2003::4 activate
SwitchC(config-router-af)#exit-address-family
SwitchC(config-router-bgp)#exit

SwitchD configuration as follows:
SwitchD(config)#router bgp 200
SwitchA(config-router)#bgp router-id 2.2.2.2
SwitchD(config-router)#neighbor 2003::3 remote-as 200
SwitchD(config-router)#neighbor 2002::2 remote-as 200
SwitchD(config-router)#address-family IPv6 unicast
SwitchD(config-router-af)#neighbor 2002::2 activate
SwitchD(config-router-af)#neighbor 2003::3 activate
SwitchD(config-router-af)#exit-address-family
SwitchD(config-router)#exit

Here the connection between SwitchB and SwitchA is EBGP, and the connection between SwitchC and SwitchD is IBGP. The BGP connection can be processed between SwitchB and SwitchD without physical link, but the premise is a route which reaches from one switch to the other switch. The route can be obtained by static routing or IGP.

14.6.4 MBGP4+ Troubleshooting

It is the same as corresponding section of BGP.

14.7 IPv6 Black Hole Routing

14.7.1 Introduction to Black Hole Routing

Black Hole Routing is a special kind of static routing which drops all the datagrams that match the routing rule.

14.7.2 IPv6 Black Hole Routing Configuration Task

1. Enable the IPv6 function
2. Configure the IPv6 Black Hole Routing

1. Enable the IPv6 function

Command	Explaination

Global Configuration Mode	
ipv6 enable	To enable the IPv6 function on the switch.

2. Configure IPv6 Black Hole Routing

Command	Explaination
Global Configuration Mode	
ipv6 route ***<ipv6-prefix/prefix-length>*** **null0 [<*precedence*>]** **no ipv6 route <*ipv6-prefix/prefix-length*> null0**	To configure static IPv6 Black Hole Routing. The no form of this command will remove the specified configuration.

14.7.3 Black Hole Routing Configuration Exmaples

Example 1: IPv6 Black Hole Routing function.

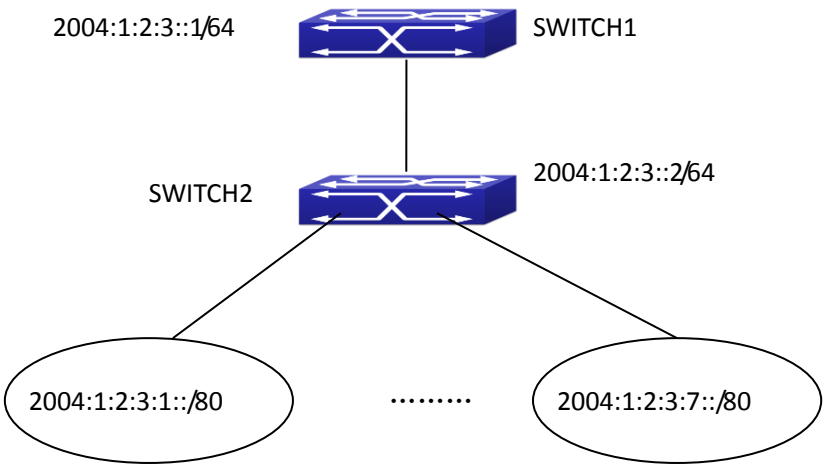


Figure 14-10 IPv6 Black Hole Routing Configuration Example

As it is shown in the figure, in Switch 2, eight in all interfaces are configured as Layer 3 VLAN interfaces for access interfaces. The network addresses are 2004:1:2:3:1/80~2004:1:2:3:7/80. A default routing is configured on Switch 2 to connect to Switch 1. And a backward default routing is configured on Switch 1 to Switch 2, whose network address is 2004:1:2:3::/64. Commonly, this configuration will work well. However, if one of the Layer 3 interfaces in Switch 2 goes down, for example, the interface belonged to 2004:1:2:3:1/80. When datagrams arrives at VLAN1 in Switch 2, there will be no routing rules for these datagrams. The switch then will forward these datagrams according to the default routing, back to Switch 1. When Switch 1 receives these datagrams, it will forward them back to Switch 2. Thus, loopback exists. To solve this problem, Black Hole Routing can be introduced on Switch 2.

ipv6 route 2004:1:2:3::/64 null0 50

Then Switch 2 will drop the datagrams from interface VLAN1 that match the Black Hole Routing rule. And loopback routing is prevented.

Configuration steps are listed as below:

Switch#config
Switch(config)#ipv6 route 2004:1:2:3::/64 null0 50

14.7.4 Black Hole Routing Troubleshooting

When configuring the Black Hole Routing function, the configuration may not work due to some reasons such as incorrect network address mask, and incorrect management distance. Attention should be paid to the following items:

- ☞ IPv6 should be enabled before IPv6 Black Hole Routing can work.
- ☞ It is suggested that the length of the network address mask should be longer than that of normal routing configuration, in order to prevent the Black Hole Routing from intervening other routing configuration.
- ☞ When the network address mask of Black Hole Routing configuration is the same with some other configuration, it is suggested that the distance of Black Hole Routing is set lower.

For problems that cannot be fixed through above methods, please issue the command show ip route distance and show ip route fib, and show l3. And copy and paste the output of the commands, and send to the technical service center of our company.

14.8 IPv6 Multicast Protocol

14.8.1 PIM-DM6

14.8.1.1 Introduction to PIM-DM6

PIM-DM6（Protocol Independent Multicast, Dense Mode）is the IPv6 version of Protocol Independent Multicast Dense Mode. It is a Multicast Routing Protocol in dense mode which adapted to small network. The members of multicast group are relatively dense under this kind of network environment. There is no difference compared with the IPv4 version PIM-DM except that the addresses it uses are IPv6 addresses. Thus we don't differentiate between PIM-DM and PIM-DM6 in this chapter. All PIM-DM in the text without specific explanation refers to IPv6 version PIM-DM.

As a result of continuous development of IPv6 network, it has the network environment of nonsupport IPv6 multicast sometimes, so it needs to do the IPv6 multicast operation by tunnel. Therefore, our PIM-DM6 supports configuration on configure tunnel, and passes through nonsupport IPv6 multicast network by single cast packet of IPv4 encapsulation.

The working process of PIM-DM can be summarized as: Neighbor Discovery, Flooding-Prune, and Graft.

1. Neigh hour Discovery

When PIM-DM router is started at beginning, Hello message is required to discover neighbors. The network nodes running PIM-DM use Hello message to contact each other. PIM-DM Hello message is sent periodically.

2. Flooding-Prune

PIM-DM assumes that all hosts on the network are ready to receive multicast data. When certain multicast source S begins to send data to a multicast group G, after receiving the multicast packet, the router will make RPF examination first according to the unicast table. If the check passes, the router will create a (S, G) table item and forward the multicast packet to all downstream PIM-DM nodes (Flooding). If the RPF examination fails, i.e. the multicast packet is inputted from the incorrect interface, and then the message is discarded. After this procedure, every node will create an (S, G) item in the PIM-DM multicast domain. If there is no multicast group member in the downstream nodes, then a Prune message is sent to upstream nodes notifying not to forward data to this multicast group any more. After receiving Prune message, the corresponding interfaces will be deleted from the output interface list corresponding with the multicast-forwarding item (S, G). Through this process, a SPT (Shortest Path Tree) is established with source S as root. Prune process is started by a sub-router.

The process above is called Flooding-Prune process. Each pruned node also provides overtime mechanism at the same time. In case of overtime of prune, the router will restart flooding-prune process. Flooding-prune of PIM-DM is conducted periodically

3. RPF examination

Adopting RPF examination, PIM-DM establishes a multicast forwarding tree initiating from data source, using existing unicast routing table. When a multicast packet arrives, the router will determine the correctness of its coming path first. If the arrival interface is the interface connected to multicast source indicated by unicast routing, then this multicast packet is considered to be from the correct path; otherwise the multicast packet will be discarded as redundant message. The unicast routing message used as path judgment can root in any Unicast Routing Protocol, such as messages found by RIP, OSPF, etc. It doesn't rely on any specific unicast routing protocol.

4. Assert Mechanism

If two multicast router A and B in the same LAN segment have their own receiving paths to multicast source S, they will respectively forward multicast data packet to LAN after receiving the packet from multicast source S. Then downstream nodes multicast router C will receive two multicast packets that are exactly the same. Once router detects such circumstance, a unique forwarder will be selected through "assert" mechanism. The optimized forwarding path is selected through "assert" packet. If the priority and costs of two or more than two paths are same, the node with a larger IP address will be selected as the upstream neighbor of item (S, G), which will be responsible for forwarding the (S, G) multicast packet.

5. Graft

When the pruned downstream node needs to recover to forwarding status, this node uses Graft Message to notify upstream nodes to resume multicast data forwarding.

14.8.1.2 PIM-DM6 Configuration Task List

1. Enable PIM-DM (Required)
2. Configure static multicast routing entries (Optional)
3. Configure additional PIM-DM parameters (Optional)

(1) Configure parameters for PIM-DM interfaces

1) Configure the interval for PIM-DM hello messages

2) Configure the interval for PIM-DM state-refresh messages

3) Configure the boundary interfaces

4) Configure the management boundary

4. Disable PIM-DM protocol

1. Enable the PIM-DM protocol

On the switch, PIM-DM can be enabled through two steps. Firstly PIM multicast routing should be enabled in global configuration mode, then PIM-DM should be configured for the specific interfaces.

Command	Explanation
Command configuration mode	
ipv6 pim multicast-routing	To enable PIM-DM multicast routing global. However, in order to enable PIM-DM for specific interfaces, the following command must be issued.

Enable PIM-SM for the specific interface:

Command	Explanation
Interface configuration mode	
ipv6 pim dense-mode	To enable PIM-DM for the specified interface (required).

2. Configure static multicast routing entries

Command	Explanation
Global configuration mode	
ipv6 mroute <*X:X::X:X*> <*X:X::X:X*> <*ifname*> <*.ifname*> **no ipv6 mroute <*X:X::X:X*> <*X:X::X:X*> [<*ifname*> <*.ifname*>]**	To configure IPv6 static multicast routing entries. The no form of this command will remove the specified routing entry.

3. Configure additional PIM-DM parameters

（**1**）Configure parameters for PIM-DM interfaces

1) Configure the interval for PIM-DM hello messages

Command	Explanation
Interface Configuration Mode	
ipv6 pim hello-interval <*interval*> **no ipv6 pim hello-interval**	To configure the interval for PIM-DM hello messages. The no form of this command will restore the default value.

2) Configure the interval for PIM-DM state-refresh messages

Command	Explanation
Interface Configuration Mode	

ipv6 pim state-refresh origination-interval **no ipv6 pim state-refresh origination-interval**	To configure the interval for sending PIM-DM state-refresh packets. The no form of this command will restore the default value.

3) Configure the boundary interfaces

Command	Explanation
Interface Configuration Mode	
ipv6 pim bsr-border **no ipv6 pim bsr-border**	To configure the interface as the boundary of PIM-DM6 protocol. On the boundary interface, STATE REFRESH messages will not be sent or received. The network connected the interface is considered as directly connected network. The no form of this command will remove the configuration.

4) Configure the management boundary

Command	Explanation
Interface Configuration Mode	
ipv6 pim scope-border <*500-599*>|<*acl_name*> **no ipv6 pim scope-border**	To configure PIM-DM6 management boundary for the interface and apply ACL for the management boundary. With default settings, ffx0::/13 is considered as the scope of the management group. If ACL is configured, then the scope specified by ACL permit command is the scope of the management group. acl_name should be standard IPv6 ACL name. The no form of this command will remove the configuration.

4. Disable PIM-DM protocol

Command	Notes
Interface Configuration Mode	
no ipv6 pim dense-mode	To disable PIM-DM for the specified interface.
Global Configuration Mode	
no ipv6 pim multicast-routing	To disable PIM-DM globally.

14.8.1.3 PIM-DM6 Typical Application

As shown in the following figure, add the Ethernet interfaces of Switch A and Switch B to corresponding vlan, and start PIM-DM Protocol on each vlan interface.

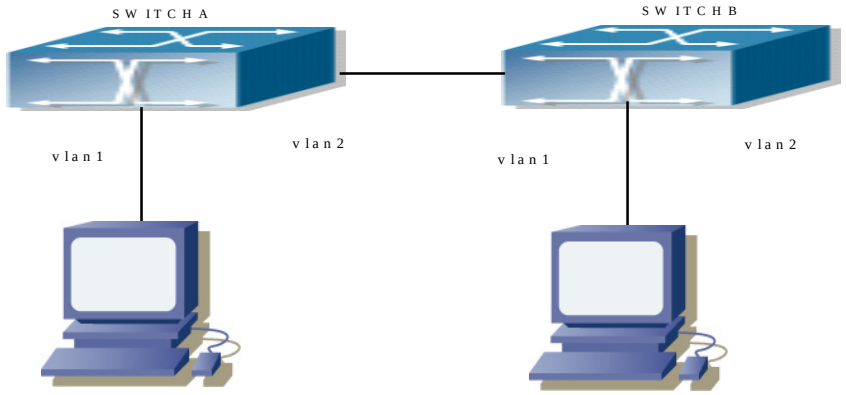


Figure 14-11 PIM-DM Typical Environment

The configuration procedure for SwitchA and SwitchB is as below:

(1) Configure SwitchA:

Switch(config)#ipv6 pim multicast-routing

Switch(config)#interface vlan 1

Switch(Config-if-Vlan1)#ipv6 address 2000:10:1:1::1/64

Switch(Config-if-Vlan1)#ipv6 pim dense-mode

Switch(Config-if-Vlan1)#exit

Switch(config)#interface vlan2

Switch(Config-if-Vlan2)#ipv6 address 2000:12:1:1:: 1/64

Switch(Config-if-Vlan2)#ipv6 pim dense-mode

(2) Configure SwitchB:

Switch(config)#ip pim multicast-routing

Switch(config)#interface vlan 1

Switch(Config-if-Vlan1)#ipv6 address 2000:12:1:1::2/64

Switch(Config-if-Vlan1)#ipv6 pim dense-mode

Switch(Config-if-Vlan1)#exit

Switch(config)#interface vlan 2

Switch(Config-if-Vlan2)#ipv6 address 2000:20:1:1::1/64

Switch(Config-if-Vlan2)#ipv6 pim dense-mode

14.8.1.4 PIM-DM6 Troubleshooting

When configuring and using PIM-DM protocol, PIM-DM protocol may fail to work normally due to physical connections, incorrect configuration and so on. So, users shall note the following points:

- Assure the physical connection is correct.
- Assure the Protocol of Interface and Link is UP (use show interface command);
- Assure PIM Protocol is turned on in Global Mode (use ipv6 pim multicast-routing command)
- Start PIM-DM Protocol on the interface (use ipv6 pim dense-mode command)

Unicast route shall be used to carry out RPF examination for multicast protocol. So the correctness of unicast route shall be guaranteed above all. If all attempts fail to solve the

problems on PIM-DM, then use debug commands such as debug ipv6 pim, copy DEBUG information in 3 minutes and send to Technology Service Center.

14.8.2 PIM-SM6

14.8.2.1 Introduction to PIM-SM6

PIM-SM6（Protocol Independent Multicast, Sparse Mode）is the IPv6 version of Protocol Independent Multicast Sparse Mode. It is a multicast routing protocol in sparse mode and mainly used in large network with group members distributed relatively sparse and wide. It is no difference from the IPv4 version PIM-SM except the addresses it uses are IPv6 addresses. Thus we don't differentiate between PIM-SM and PIM-SM6 in this chapter. All PIM-SM in the text without specific explanation is IPv6 version PIM-SM. Unlike the Flooding-Prune of Dense Mode, PIM-SM Protocol assumes no host needs receiving multicast data packets. PIM-SM router forwards multicast data packets to a host only on definite request.

By setting RP (Rendezvous Point) and BSR (Bootstrap Router), PIM-SM announce multicast packet to all PIM-SM routers and establish, using Join/Prune message of routers, RPT (RP-rooted shared tree) based on RP. Consequently the network bandwidth occupied by data packets and control messages is cut down and the transaction cost of routers is reduced. Multicast data get to the network segment where the multicast group members are located along the shared tree flow. When the data traffic reaches a certain amount, multicast data stream can be switched to source-based SPT (Shortest Path Tree) to shorten network delay. PIM-SM doesn't rely on any specific unicast routing protocol but make RPF examination using existing unicast routing table.

1. PIM-SM Working Principle

The working process of PIM-SM mainly includes neighbor discovery, creation of RPT, registration of multicast source, SPT switch and so on. The neighbor discovery mechanism is the same with the mechanism of PIM-DM. We won't introduce any more.

(1) Creation of RP Shared Tree (RPT)

When a host joins a multicast group G, the leaf router directly connected with the host finds out through IGMP message that there is a receiver of multicast group G, then it works out the corresponding Rendezvous Point RP for multicast group G, and send join message to upper level nodes in RP direction. Every router on the way from the leaf router to RP will create a (*, G) table item, indicating the message from any source to multicast group G is suitable for this item. When RP receives the message sent to multicast group G, the message will get to the leaf router along the established path and then reach the host. In this way, the RPT with RP as root is created.

(2) Multicast Source Registration

When multicast source S sends a multicast packet to multicast group G, the PIM-SM multicast router directly connected to it will take charge of sealing the multicast packet into registered message and unicast it to corresponding RP. If there are more than one PIM-SM multicast routers on a network segment, then DR (Designated Router) takes charge of forwarding

the multicast packet.

(3) SPT Switch

Once the multicast router finds that the rate of the multicast packet from RP with destination address G exceeds threshold, the multicast router will send Join message to the upper lever nodes in the source direction, which results in the switch from RPT to SPT.

2. Preparation before PIM-SM configuration

(1) Configuration Candidate RP

More than one RPs (candidate RP) are permitted in PIM-SM network and each C-RP (Candidate RP) takes charge of forwarding multicast packets with destination address in a certain range. To configure more than one candidate RPs can achieve RP load balancing. There is no master or slave difference among RPs. All multicast routers work out the RP corresponded with certain multicast group based on the same algorithm after receiving the candidate RP message announced by BSR.

Note that one RP can serve more than one multicast groups, even all multicast groups. But each multicast group can only correspond with one unique RP at any moment. It can't correspond with more RPs at the same time.

(2) BSR Configuration

As the management core of PIMSM network, BSR is in charge of collecting messages sent by candidate RPs and broadcast them..

There may be only one BSR within a network. However, there may be several candidate BSRs to be configured. With such arrangement, once a BSR fails, another may be switched to. C-BSR determines BSR through automatic selection.

14.8.2.2 PIM-SM6 Configuration Task List

1. Enable PIM-SM (Required)
2. Configure static multicast routing entries (Optional)
3. Configure additional parameters for PIM-SM (Optional)

(1) Configure parameters for PIM-SM interfaces

1) Configure the interval for PIM-SM hello messages
2) Configure the holdtime for PIM-SM hello messages
3) Configure ACL for PIM-SM6 neighbors
4) Configure the interface as the boundary interface of the PIM-SM6 protocol
5) Configure the interface as the management boundary of the PIM-SM6 protocol

(2) Configure global PIM-SM parameters

1) Configure the switch as a candidate BSR
2) Configure the switch as a candidate RP

5)Configure static RP

3) Configure the cache time of kernel multicast route

4. Disable the PIM-SM protocol

1. Enable PIM-SM protocol

The PIM-SM protocol can be enabled on Layer 3 switches by enabling PIM6 in global configuration mode and then enabling PIM-SM for specific interfaces in the interface configuration mode.

Command	Explanation
Global Configuration Mode	
[no] ipv6 pim multicast-routing	To enable the PIM-SM6 protocol for all the interfaces (However, in order to make PIM-SM work for specific interfaces, the following command should be issued). (required)

Make the PIM-SM protocol work for specific interfaces

Command	Explanation
Interface Configuration Mode	
[no] ipv6 pim sparse-mode [passive]	To enable PIM-SM for the specified interface. The no form of this command will disable the PIM-SM protocol (required).

2. Configure static multicast routing entries

Command	Explanation
Global Configuration Mode	
ipv6 mroute <*X:X::X:X*> <*X:X::X:X*> <*ifname*> <.*ifname*> **no ipv6 mroute <*X:X::X:X*> <*X:X::X:X*> [<*ifname*> <.*ifname*>]**	To configure a static multicast routing entry. The no form of this command will remove the specified static multicast routing entry.

3. Configure the additional parameters for PIM-SM

（1）Configure parameters for PIM-SM interfaces

1) Configure the interval for PIM-SM hello messages

Command	Explanation
Interface Configuration Mode	
ipv6 pim hello-interval <*interval*> **no ipv6 pim hello-interval**	To configure the interval for PIM-SM hello messages. The no form of this command restores the interval to the default value.

2) Configure the hold time for PIM-SM6 hello messages

Command	Explanation
Interface Configuration Mode	
ipv6 pim hello-holdtime <*value*> **no ipv6 pim hello-holdtime**	To configure the value of the holdtime field in the PIM-SM hello messages. The no form of this command will restore the hold time to the default value.

3) Configure ACL for PIM-SM6 neighbors

Command	Explanation
Interface Configuration Mode	

ipv6 pim neighbor-filter <*access-list-name*> **no ipv6 pim neighbor-filter <*access-list-name*>**	To configure ACL to filter PIM-SM6 neighbor. If session to the neighbor has been denied by ACL, then the sessions that have been set up will be discarded immediately and new sessions will not be set up.

4) Configure the interface as the boundary interface of the PIM-SM6 protocol

Command	Explanation
Interface Configuration Mode	
ipv6 pim bsr-border **no ipv6 pim bsr-border**	To configure the interface as the boundary of PIM-SM6 protocol. On the boundary interface, BSR messages will not be sent or received. The network connected the interface is considered as directly connected network. The no form of this command will remove the configuration.

5) Configure the interface as the management boundary of the PIM-SM6 protocol

Command	Explanation
Interface Configuration Mode	
ipv6 pim scope-border <*500-599*>|<*acl_name*> **no ipv6 pim scope-border**	To configure PIM-SM6 management boundary for the interface and apply ACL for the management boundary. With default settings, ffx0::/13 is considered as the scope of the management group. If ACL is configured, then the scope specified by ACL permit command is the scope of the management group. acl_name should be standard IPv6 ACL name. The no form of this command will remove the configuration.

（2）Configure global PIM-SM6 parameter

1) Configure the switch as a candidate BSR

Command	Explanation
Global Configuration Mode	
**ipv6 pim bsr-candiate {vlan <*vlan_id*>|<*ifname*>	tunnel <*1-50*>}[hash-mask-length] [priority]** **no Ipv6 pim bsr-candiate {vlan <*vlan_id*>|<*ifname*>	tunnel <*1-50*>}[hash-mask-length] [priority]**	This command is the global candidate BSR configuration command, which is used to configure the information of PIM-SM candidate BSR so that it can compete for BSR router with other candidate BSR. The no operation is to cancel the configuration of BSR.

2) Configure the switch as a candidate RP

Command	Explanation
Global Configuration Mode	

ipv6 pim rp-candiate {vlan<*vlan-id*>| loopback<*index*>|<*ifname*>} [<*group range*>] [<*priority*>] **no ipv6 pim rp-candiate**	This command is the global candidate RP configuration command, which is used to configure the information of PIM-SM candidate RP so that it can compete for RP router with other candidate RP. The no operation is to cancel the configuration of RP.

3) Configure static RP

Command	Explanation
Global Configuration Mode	
ipv6 pim rp-address <*rp-address*> [<*group-range*>] **no ipv6 pim rp-address <*rp-address*> {all|<*group-range*>}**	To configure the address of the candidate RP. The no form of this command will remove the configuration for the candidate RP.

4) Configure the cache time of kernel multicast route

Command	Explanation
Global Configuration Mode	
ipv6 multicast unresolved-cache aging-time <value> **no ipv6 multicast unresolved-cache aging-time**	Configure the cache time of kernel multicast route, the no command restores the default value.

4. Disable PIM-SM protocol

Command	Explanation
Interface Configuration Mode	
no ipv6 pim sparse-mode	To disable the PIM-SM6 protocol.
Global Configuration Mode	
no ipv6 pim sparse-mode	To disable PIM-DM globally.

14.8.2.3 PIM-SM6 Typical Application

As shown in the following figure, add the Ethernet interfaces of SwitchA, SwitchB, SwitchC and SwitchD to corresponding VLAN, and start PIM-SM Protocol on each VLAN interface.

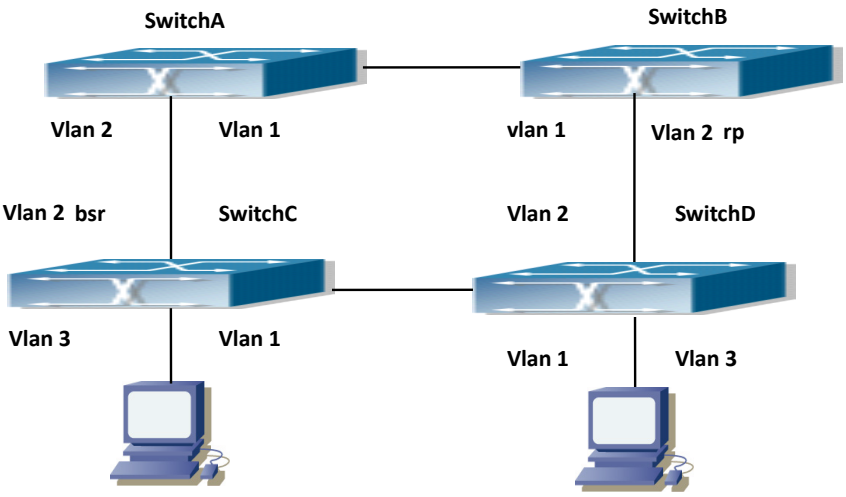


Figure 14-22 PIM-SM Typical Environment

The configuration procedure for SwitchA, SwitchB, SwitchC and SwitchD is as below:

(1) Configure SwitchA:

Switch(config)#ipv6 pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ipv6 address 2000:12:1:1::1/64
Switch(Config-if-Vlan1)#ipv6 pim sparse-mode
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)#ipv6 address 2000:13:1:1::1/64
Switch(Config-if-Vlan2)#ipv6 pim sparse-mode

(2) Configure Switch B:

Switch(config)#ipv6 pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ipv6 address 2000:12:1:1::2/64
Switch(Config-if-Vlan1)#ipv6 pim sparse-mode
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)#ipv6 address2000:24:1:1::2/64
Switch(Config-if-Vlan2)#ipv6 pim sparse-mode
Switch(Config-if-Vlan2)#exit
Switch(config)#ipv6 pim rp-candidate vlan2

(3) Configure SwitchC:

Switch(config)#ipv6 pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ipv6 address 2000:34:1:1::3/64
Switch(Config-if-Vlan1)#ipv6 pim sparse-mode
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)#ipv6 address 2000:13:1:1::3/64
Switch(Config-if-Vlan2)#ipv6 pim sparse-mode

Switch(Config-if-Vlan2)#exit
Switch(config)#interface vlan 3
Switch(Config-if-Vlan3)#ipv6 address 2000:30:1:1::1/64
Switch(Config-if-Vlan3)#ipv6 pim sparse-mode
Switch(Config-if-Vlan3)#exit
Switch(config)#ipv6 pim bsr-candidate vlan2 30 10

(4) Configure SwitchD:
Switch(config)#ipv6 pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-if-Vlan1)#ipv6 address 2000:34:1:1::4/64
Switch(Config-if-Vlan1)#ipv6 pim sparse-mode
Switch(Config-if-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-if-Vlan2)#ipv6 address 2000:24:1:1::4/64
Switch(Config-if-Vlan2)#ipv6 pim sparse-mode
Switch(Config-if-Vlan2)#exit
Switch(config)#interface vlan 3
Switch(Config-if-Vlan3)#ipv6 address 2000:40:1:1::1/64
Switch(Config-if-Vlan3)#ipv6 pim sparse-mode

14.8.2.4 PIM-SM6 Troubleshooting

When configuring and using PIM-SM protocol, PIM-SM protocol may fail to work normally due to physical connections, incorrect configuration and so on. So, users shall note the following points:

- Assure the physical connection is correct.
- Assure the Protocol of Interface and Link is UP (use show interface command);
- Unicast route shall be used to carry out RPF examination for multicast protocol. So the correctness of unicast route shall be guaranteed above all.
- PIM-SM Protocol requires supports of RP and BSR, therefore you should use show ipv6 pim bsr-router first to see if there is BSR information. If not, you need to check if there is unicast routing leading to BSR.
- Use show ipv6 pim rp-hash command to check if RP information is correct; if there is no RP information, you still need to check unicast routing;

If all attempts fail to solve the problems on PIM-SM, then use debug commands such as debug ipv6 pim/ debug ipv6 pim bsr, copy DEBUG information in 3 minutes and send to Technology Service Center.

14.8.3 ANYCAST RP v6

14.8.3.1 Introduction to ANYCAST RP v6

Anycast RP v6 is a technology based on PIM protocol, which provides redundancy in order to recover as soon as possible once an RP becomes unusable.

The kernel concept of Anycast RP v6 is that the RP addresses configured all over the whole network exist on multiple multicast servers (the most common situation is that every device providing ANYCAST RP uses LOOPBACK interface, and using the longest mask to configures RP addresses on this interface), while the unicast routing algorithm will make sure that PIM routers can always find the nearest RP, thus , providing a shorter and faster way to find RP in a larger network., Once an RP being used becomes unusable, the unicast routing algorithm will ensure that the PIM router can find a new RP path fast enough to recover the multicast server in time. Multiple RP will cause a new problem that is if the multicast source and the receivers are registered to different RP, some receivers will not be able to receive data of multicast source (obviously, the register messages only prefer the nearest RP). So, in order to keep the communication between all RP, Anycast RP defines that the nearest RP to the multicast source should forward the source register messages to all the other RP to guarantee that all joiners of the RP can find the multicast source.

The method to realize the PIM-protocol-based Anycast RP is that: maintaining an ANYCAST RP list on every switch configured with Anycast RP and using another address as the label to identify each other. When one Anycast RP device receives a register message, it will send the register message to other Anycast RP devices while using its own address as the source address, to notify all the other devices of the original destination.

14.8.3.2 ANYCAST RP v6 Configuration Task

1. Enable ANYCAST RP v6 function
2. Configure ANYCAST RP v6

1. Enable ANYCAST RP v6 function

Command	Explanation
Global Configuration Mode	
ipv6 pim anycast-rp **no ipv6 pim anycast-rp**	Enable ANYCAST RP function. (necessary) The no operation will globally disable the ANYCAST RP function.

2. Configure ANYCAST RP v6

(1) Configure RP candidate

Command	Explanation
Global Configuration Mode	
ipv6 pim rp-candidate {vlan<*vlan-id*> |loopback<*index*> |<*ifname*>} [<*A:B::C:D*>][<*priority*>] **no ipv6 pim rp-candidate**	Now, the PIM-SM has allowed the Loopback interface to be a RP candidate.(necessary) Please pay attention to that, ANYCAST RP protocol can configure the Loopback interface

	or a regular three-layer VLAN interface to be the RP candidate. In make sure that PIM routers in the network can find where the RP locates, the RP candidate interface should be added into the router. No operation will cancel the RP candidate configured on this router.

(2) Configure self-rp-address (the RP communication address of this router)

Command	Explanation
Global Configuration Mode	
ipv6 pim anycast-rp self-rp-address A:B::C:D **no ipv6 pim anycast-rp self-rp-address**	Configure the self-rp-address of this router (as a RP). This address can be used to exclusively identify this router when communicating with other RP.(necessary) the effect of **self-rp-address** refers to two respects: 1 Once this router (as a RP) receives the register message from a DR unicast, it needs to forward the register message to all the other RP in the network, notifying them of the state of source (S.G). While forwarding the register message, this router will change the source address of it into self-rp-address. 2 Once this router(as a RP) receives a register message from other RP unicast, such as a register message whose destination is the self-rp-address of this router, it will create (S,G) state and send back a register-terminating message, whose destination address is the source address of the register message. Pay attention: self-rp-address has to be the address of a three-layer interface on this router, but the configuration is allowed to be done with the absence of the interface. The self-rp-address should be unique. No operation will cancel the self-rp-address which is used to communicate with other RP by this router.

(3) Configure other-rp-address (other RP communication addresses)

Command	Explanation
Global Configuration Mode	
ipv6 pim anycast-rp <*anycast-rp-addr*>	Configure anycast-rp-addr on this router (as a

<other-rp-addr> **no ipv6 pim anycast-rp <*anycast-rp-addr*> <*other-rp-addr*>**	RP). This unicast address is actually the RP address configured on multiple RP in the network, in accordance with the address of RP candidate interface (or Loopback interface). The effect of **anycast-rp-addr** includes: 1 Although more than one anycast-rp-addr addresses are allowed to be configured, only the one having the same address with the currently configured RP candidate address will take effect. Only after that, can the other-rp-address in accordance with this anycast-rp-addr take effect. 2 The configuration is allowed to be done with the absence of the interface in accordance with the anycast-rp-addr. Configure on this router (as a RP) the other-rp-addresses of other RP communicating with it. This unicast address identifies other RP and is used in the communication with local routers. The effect of **other-rp-address** refers to two respects: 1 Once this router (as a RP) receives the register message from a DR unicast, it should forward it to other RP in the network to notify all the RP in the network of the source (S.G) state. While forwarding, the router will change the destination address of the register message into other-rp-address. 2 Multiple other-rp-addresses can be configured in accordance with one anycast-rp-addr, Once the register message from a DR is received, it should be forwarded to all of this RP one by one. No operation will cancel other-rp-address communicating with this router.

14.8.3.3 ANYCAST RP v6 Configuration Examples

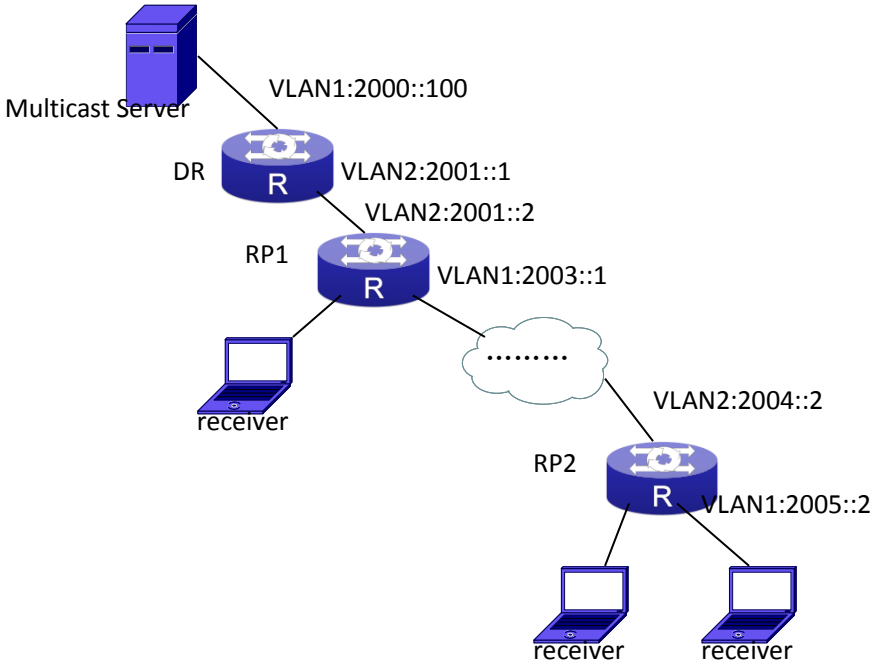


Figure 14-13 The ANYCAST RP v6 function of a router

The following is the configuration steps:

RP1 Configuration:

Switch#config
Switch(config)#interface loopback 1
Switch(Config-if-Loopback1)#ipv6 address 2006::1/128
Switch(Config-if-Loopback1)#exit
Switch(config)#ipv6 pim rp-candidate loopback1
Switch(config)#ipv6 pim bsr-candidate vlan 1
Switch(config)#ipv6 pim multicast-routing
Switch(config)#ipv6 pim anycast-rp
Switch(config)#ipv6 pim anycast-rp self-rp-address 2003::1
Switch(config)#ipv6 pim anycast-rp 2006::1 2004::2

RP2 Configuration:

Switch#config
Switch(config)#interface loopback 1
Switch(Config-if-Loopback1)#ipv6 address 2006::1/128
Switch(Config-if-Loopback1)#exit
Switch(config)#ipv6 pim rp-candidate loopback1
Switch(config)#ipv6 pim multicast-routing
Switch(config)#ipv6 pim anycast-rp
Switch(config)#ipv6 pim anycast-rp self-rp-address 2004::2
Switch(config)#ipv6 pim anycast-rp 2006::1 2003::1

Please pay attention to that, for promulgating loopback interface router, if use MBGP4+ protocol, then can use network command; or use RIPng protocol, then can use route command.

14.8.3.4 ANYCAST RP v6 Troubleshooting

When configuring and using ANYCAST RP v6 function, the ANYCAST RP might work abnormally because of faults in physical connections, configurations or something others. So, the users should pay attention to the following points:

- The physical connections should be guaranteed to be correct
- The PIM-SM6 protocol should be guaranteed to operate normally
- The ANYCAST RP should be guaranteed to be enabled in Global configuration mode
- The self-rp-address should be guaranteed to be configured correctly in Global configuration mode
- The other-rp-address should be guaranteed to be configured correctly in Global configuration mode
- All the interface routers should be guaranteed to be correctly added, including the loopback interface as a RP
- Use "**show ipv6 pim anycast rp status**" command to check whether the configuration information of ANYCAST RP is correct

If the problems of ANYCAST still cannot be solved after checking, please use debug commands like "debug ipv6 pim anycast-rp", then copy the DEBUG information within three minutes and send it to the technical service center of our company.

14.8.4 PIM-SSM6

14.8.4.1 Introduction to PIM-SSM6

Source Specific Multicast (PIM-SSM6) is a new kind of multicast service protocol. With PIM-SSM6, a multicast session is distinguished by the multicast group address and multicast source address. In SSM6, hosts can be added into the multicast group manually and efficiently like the traditional PIM-SM6, but leave out the shared tree and RP management in PIM-S6M. In SSM6, SPT tree will be constructed with (S,G). G for the multicast group address and S for the source address of the multicast which sends datagram to G. (S,G) in a pair is named as a channel of SSM6. SSM6 serves best for the application of multicast service which is from one station to many ones, for example, the network sports video channel, and the news channel. By default, the multicast group address of SSM6 is limited to ff3x::/32. However this address range can be extended according to actual situations.

PIM-SSM6 can be supported in the PIM-DM6 environment.

14.8.4.2 PIM-SSM6 Configuration Task List

Command	Explanation
Global configuration mode	

ipv6 pim ssm {default|range *<access-list-number>*} **no ipv6 pim ssm**	To configure address range for pim-ssm multicast group. The no prefix will disable this command.

14.8.4.3 PIM-SSM6 Configuration Example

As it is shown in the below figure, ethernet interfaces of switchA, switchB, switchC, and switchD are separated into different vlan. And PIM-SM6 or PIM-DM6 is enabled on all the vlan interfaces. Take configuration of PIM-SM6 for example.

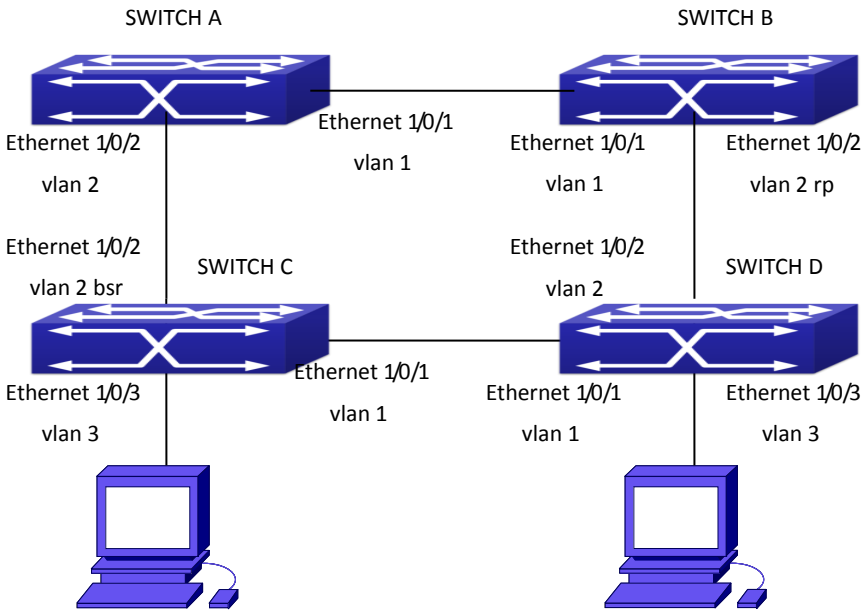


Figure 14-14 PIM-SSM typical environment

Configurations of switchA , switchB, switchC and switchD are listed as below:

(1) Configuration of switchA：

Switch(config)#ipv6 pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-If-Vlan1)# ipv6 address 2000:12:1:1::1/64
Switch(Config-If-Vlan1)# ipv6 pim sparse-mode
Switch(Config-If-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-If-Vlan2)# ipv6 address 2000:13:1:1::1/64
Switch(Config-If-Vlan2)# ipv6 pim sparse-mode
Switch(Config-If-Vlan2)#exit
Switch(config)#ipv6 access-list 500 permit ff1e::1/64
Switch(config)#ip pim ssm range 500

(2)Configuration of switchB：

Switch(config)#ipv6 pim multicast-routing
Switch(config)#interface vlan 1

```
Switch(Config-If-Vlan1)# ipv6 address 2000:12:1:1::2/64
Switch(Config-If-Vlan1)# ipv6 pim sparse-mode
Switch(Config-If-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-If-Vlan2)# ipv6 address2000:24:1:1::2/64
Switch(Config-If-Vlan2)# ipv6 pim sparse-mode
Switch(Config-If-Vlan2)# exit
Switch(config)# ipv6 pim rp-candidate vlan2
Switch(config)#ipv6 access-list 500 permit ff1e::1/64
Switch(config)#ip pim ssm range 500
```

(3) Configuration of SwitchC：

```
Switch(config)#ipv6 pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-If-Vlan1)# ipv6 address 2000:34:1:1::3/64
Switch(Config-If-Vlan1)# ipv6 pim sparse-mode
Switch(Config-If-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-If-Vlan2)# ipv6 address 2000:13:1:1::3/64
Switch(Config-If-Vlan2)# ipv6 pim sparse-mode
Switch(Config-If-Vlan2)#exit
Switch(config)#interface vlan 3
Switch(Config-If-Vlan3)# ipv6 address 2000:30:1:1::1/64
Switch(Config-If-Vlan3)# ipv6 pim sparse-mode
Switch(Config-If-Vlan3)# exit
Switch(config)# ipv6 pim bsr-candidate vlan2 30 10
Switch(config)#ipv6 access-list 500 permit ff1e::1/64
Switch(config)#ip pim ssm range 500
```

(4) Configuration of SwitchD：

```
Switch(config)#ipv6 pim multicast-routing
Switch(config)#interface vlan 1
Switch(Config-If-Vlan1)# ipv6 address 2000:34:1:1::4/64
Switch(Config-If-Vlan1)# ipv6 pim sparse-mode
Switch(Config-If-Vlan1)#exit
Switch(config)#interface vlan 2
Switch(Config-If-Vlan2)# ipv6 address 2000:24:1:1::4/64
Switch(Config-If-Vlan2)# ipv6 pim sparse-mode
Switch(Config-If-Vlan2)#exit
Switch(config)#interface vlan 3
Switch(Config-If-Vlan3)# ipv6 address 2000:40:1:1::1/64
Switch(Config-If-Vlan3)# ipv6 pim sparse-mode
Switch(Config-If-Vlan3)#exit
Switch(config)#ipv6 access-list 500 permit ff1e::1/64
```

Switch(config)#ip pim ssm range 500

14.8.4.4 PIM-SSM6 Troubleshooting

When configuring the PIM-SSM6 protocol, it may fail to work because of the failure of physical connection or the mis-configurations. To debug these errors, attention should be paid to the following lists.

- ☞ Make sure the physical links are connected correctly.
- ☞ Make sure the state of the data link layer has become UP. (Use show interface command).
- ☞ Make sure PIM6 is enabled in global configuration mode (Refer to the command ipv6 pim multicast-routing).
- ☞ Make sure PIM-SM6 is configured on the interface (Refer to the command ipv6 pim sparse-mode)
- ☞ Make sure SSM6 is configure in global configuration mode.
- ☞ The multicast protocol uses the unicast routing to make RPF check. Hence, single-cast routing should be verified firstly.

If problems could not be fixed with the above check list, please enable the command of **debug ipv6 pim event and debug ipv6 pim packet**, and save the debug information for 3 minutes, and send if to Technology Service Center.

14.8.5 IPv6 DCSCM

14.8.5.1 Introduction to IPv6 DCSCM

The technology of IPv6 DCSCM (Destination Control and Source Control Multicast) includes three aspects: the multicast source control, the multicast user control and the service-priority-oriented policy multicast.

IPv6 DCSCM Controllable Multicast technology proceeds as the following way:

1. If source controlled multicast is configured on the edge switches, only the multicast data of the specified group from the specified source can pass.
2. The RP switches which are the core of PIM-SM will directly send REGISTER_STOP as response to the REGISTER messages not from the specified source and specified group, and no entry is allowed to be created. (This task is implemented in the PIM-SM module).

The control of multicast users of IPv6 DCSCM technology is implemented on the basis of controlling the MLD message sent from the users, so the control module is MLD snooping and the MLD module, the control logic of which includes the following three methods: controlling according to the VLAN+MAC sending the message, controlling according to the IP address sending the message, and controlling according to the input port of the message. MLD snooping can adopts all the three methods at the same time, while the MLD module, at the third layer, can only control the IP address sending the message.

The service-priority-oriented policy multicast of IPv6 DCSCM technology adopts the following method: for the confined multicast data, the user-specified priority will be set at the access point, enabling the data can be sent at a higher priority through TRUNK, and guaranteeing that the data can be sent through the whole net at the user-specified priority.

14.8.5.2 IPv6 DCSCM Configuration Task Sequence

1. The source control configuration
2. The destination control configuration
3. The multicast policy configuration

1. The source control configuration

The source control configuration has three steps, first is globally enabling the source control, the following is the command of globally enabling the source control:

Command	Explanation
Global Configuration Mode	
ipv6 multicast source-control(necessary) **no ipv6 multicast source-control**	Globally enable the source control, the no operation of this command will globally disable the source control. What should be paid attention to is that, once globally enable the source control, all the multicast messages will be dropped by default. All the source control configurations can only be done after globally enabled, and only when all the configured rules are disabled, the source control can be disabled globally.

The next is configuring the source control rules, which adopts the same method as configuring ACL, using ACL number from 8000 to 8099, while each rule number can configure 10 rules. What should be paid attention to is that these rules have orders, the earliest configured rule is at the front. Once a rule is matched, the following ones will not take effect, so the globally enabled rules should be the last to configure. The following is the command:

Command	Explanation
Global Configuration Mode	
[no] ipv6 access-list *<8000-8099>* {deny|permit} {{<*source/M*>}|{host-source <*source-host-ip*>}|any-source} {{<*destination/M*> }|{host-destination <*destination-host-ip*>}|any-destination}	Used to configure the source control rules, the rules can only take effect when applied to the specified port. The no operation of this command can delete the specified rule.

The last is to configure the rules to the specified port.

Pay attention: since the configured rules will take up entries of hardware, configuring too many rules might cause failure if the underlying entries are full, so it is recommended that users adopt rules as simple as possible. The following is the configuration command:

Command	Explanation
Port Configuration Mode	
[no] ipv6 multicast source-control access-group <*8000-8099*>	Used to configure the source control rule to a port, he no operation will cancel this configuration.

2. The configuration of destination control

The configuration of destination control is similar to that of source control, and also has three steps:

First, globally enable the destination control, since destination control needs to avoid the unauthorized users from receiving multicast data, once it is enabled globally, the switch will stop broadcasting received multicast data, so if a switch has enabled destination control, users should not connect two or more other Layer three switches within the same VLAN where it locates. The following is the configuration command:

Command	Explanation
Global Configuration Mode	
multicast destination-control(necessary)	Globally enable IPV4 and IPv6 destination control, the no operation of this command will globally disable destination control. All of the other configuration can only take effect after globally enabled.

The next is configuring destination control rules, which are similar to that of source control, but using ACL number from 9000 to 10099 instead.

Command	Explanation
Global Configuration Mode	
[no] ipv6 access-list <*9000-10099*> {deny|permit} {{<*source/M*>}|{host-source <*source-host-ip*>}|any-source} {{<*destination/M*>}|{host-destination <*destination-host-ip*>}|any-destination}	Used to configure destination control rules, these rules can only take effect when applied to specified source IP, VLAN-MAC or port. The no operation of this rule will delete the specified rule.

The last step is to configure the rules to the specified source IP, source VLAN MAC or the specified port. What should be paid attention to is that only when the MLD-SNOOPING is enabled, these rules can be globally used, or, only rules of source IP can be used in MLD protocol. The following is the configuration command:

Command	Explanation
Port Mode	
[no] ipv6 multicast destination-control access-group <*9000-10099*>	Used to configure the destination control rule to a port, the no operation of this command will cancel the configuration.
Global Configuration Mode	

[no] ipv6 multicast destination-control <*1-4094*> <*macaddr*> access-group <*9000-10099*>	Used to configure the destination control rules to the specified VLAN-MAC, the no operation of this command will cancel the configuration.
[no] ipv6 multicast destination-control <*IPADDRESS/M*> access-group <*9000-100999*>	Used to configure the destination control rules to the specified source IPv6 address/MASK, the no operation of this command will cancel the configuration.

3. The configuration of multicast policy

The multicast policy adopts the method of specifying a priority for the specified multicast data to meet the user's particular demand, what should be paid attention to is that only when multicast data is transmitted in TRUNK, can it be taken special care of. The configuration is quite simple, for only one command is needed, that is set priority for the specified multicast, the following is the command:

Command	Explanation
Global Configuration Mode	
[no] ipv6 multicast policy <*IPADDRESS/M*> <*IPADDRESS/M*> cos <*priority*>	Configure multicast policy, set priority for sources and groups in a specified range, the priority valid range is 0 to 7.

14.8.5.3 IPv6 DCSCM Typical Examples

1. Source control

In order to prevent an edge switch sends multicast data at will, we configure on the edge switch that only the switch whose port is Ethernet1/0/4 can send multicast data, and the group of data should be ff1e::1. The uplink port Ethernet1/0/25 can forward multicast data without being restricted, so we can configure as follows.

Switch(config)#ipv6 access-list 8000 permit any-source ff1e::1
Switch(config)#ipv6 access-list 8001 permit any any
Switch(config)#ipv6 multicast source-control
Switch(config)#interface Ethernet1/0/4
Switch(Config-If-Ethernet1/0/4)#ipv6 multicast source-control access-group 8000
Switch(config)#interface Ethernet1/0/25
Switch(Config-If-Ethernet1/0/25)#ipv6 multicast source-control access-group 8001

2. Destination control

We want to confine that the users of the segment whose address is fe80::203:fff:fe01:228a/64 can not join the ff1e::1/64 group, so we can configure as follows:

First, enable MLD Snooping in the VLAN where it locates (in this example, it is VLAN2).

Switch(config)#ipv6 mld snooping
Switch(config)#ipv6 mld snooping vlan 2

Then configure relative destination control access list and configure specified IPv6 address to use this access list.

Switch(config)#ipv6 access-list 9000 deny any ff1e::1/64

Switch(config)#ipv6 access-list 9000 permit any any
Switch(config)#multicast destination-control
Switch(config)#ipv6 multicast destination-control fe80::203:fff:fe01:228a/64 access-group 9000
Thus, the users of this segment can only join groups other than 2ff1e::1/64.

3. Multicast policy

Server 2008::1 is sending important multicast data in group ff1e::1, we can configure on its access switch as follows:

Switch(config)#ipv6 multicast policy 2008::1/128 ff1e::1/128 cos 4

Thus this multicast flow will have a priority of 4, when it passes the TRUNK port of this switch to another switch (generally speaking, it is a relatively high priority, the data with higher priority might be protocol data, if a higher priority is set, when there is too much multicast data, the switch protocol might operate abnormally).

14.8.5.4 IPv6 DCSCM Troubleshooting

IPv6 DCSCM module acts like ACL, so most problems are caused by improper configuration. Please read the instructions above carefully.

14.8.6 MLD

14.8.6.1 Introduction to MLD

MLD (Multicast Listener Discovery) is the multicast group member (receiver) discovery protocol serving IPv6 multicast. It is similar to IGMP Protocol in IPv4 multicast application. Correspondingly, MLD Protocol version1 is similar to IGMP Protocol version2, and MLD Protocol version2 is similar to IGMP Protocol version3. Current firmware supports MLDv1/ MLDv2.

The IPv6 multicast hosts can join or leave from multicast group at any location, any time, regardless of the total number of group members. It is unnecessary and impossible for multicast switch to store the relationship among all host members. Multicast switch simply finds out via MLD protocol if there are receivers of certain multicast group on the network segment connected to each port. The only thing host need to do is to keep the record of which multicast groups it joined.

MLD is unsymmetrical between host and switch: the host needs to respond the MLD query message of multicast switch with membership report message; the switch periodically sends membership query message and determines if there is host joining a specific group in its subnetworks according to the response message received, and after it receives the report of a host quitting from the group, it sends out the query for the group to confirm if there is no member left in it.

There are three types of protocol messages of MLD Protocol, that is, Query, Report and Done (which is corresponding to Leave of IGMPv2). Like IGMPV2, the Query messages include General Query and Specific Group Query. General Query uses the multicast address FF02::1 of

hosts as destination address, the group address is 0; and Specific Group Query use its group address as destination address. The multicast addresses of MLD use 130, 131 and 132 as data types denoting the three kinds of messages mentioned above. Other logic is basically same as IGMPv2.

MLD protocol version2 use FF02::16 as destination address of membership report, and 143 as data type. The other logic of MLD Protocol version2 is similar to IGMP Protocol version3.

14.8.6.2 MLD Configuration Task List

1、 Start MLD (Required)

2、 Configure MLD auxiliary parameters (Required)

（1）Configure MLD group parameters

1）Configure MLD group filter conditions

（2）Configure MLD query parameters

1）Configure the interval of MLD sending query message

2）Configure the maximum response time of MLD query

3）Configure overtime of MLD query

3、 Shut down MLD Protocol

1. Start MLD Protocol

There is no special command for starting MLD Protocol on EDGECORE series layer 3 switches. MLD Protocol will automatically start up as long as any IPv6 multicast protocol is started on corresponding interface.

Command	Explanation
Global Mode	
Ipv6 pim multicast-routing	To start Global IPv6 Multicast Protocol, the precondition of starting MLD Protocol. The NO operation of corresponding command shuts ipv6 multicast protocol and MLD Protocol. (Required)

Command	**Explanation**
Port Configuration Mode	
ipv6 pim dense-mode | ipv6 pim sparse-mode	Start MLD Protocol. The NO operation of corresponding command shuts MLD Protocol. (Required)

2. Configure MLD auxiliary parameters

（1）Configure MLD group parameters

1） Configure MLD group filter conditions

Command	**Explanation**
Port Configuration Mode	
ipv6 mld access-group ***<acl_name>*** **no ipv6 mld access-group**	Configure the filter conditions of interface for MLD group; the NO operation of this command cancels filter conditions.

（2）Configure MLD Query parameters

1）Configure interval time for MLD to send query messages

2）Configure the maximum response time of MLD query

3）Configure the overtime of MLD query

Command	**Explanation**
Port Configuration Mode	
ipv6 mld query-interval <*time_val*> **no ipv6 mld query-interval**	Configure the interval of MLD query messages sent periodically; the NO operation of this command restores the default value.
ipv6 mld query-max-response-time <*time_val*> **no ipv6 mld query-max-response-time**	Configure the maximum response time of the interface for MLD query; the NO operation of this command restores the default value.
ipv6 mld query-timeout <*time_val*> **no ipv6 mld query-timeout**	Configure the overtime of the interface for MLD query; the NO operation of this command restores the default value.

3. Shut down MLD Protocol

Command	**Explanation**
Port Configuration Mode	
no ipv6 pim dense-mode | no ipv6 pim sparse-mode | no ipv6 pim multicast-routing (Global Mode)	Shut down MLD Protocol

14.8.6.3 MLD Typical Application

As shown in the following figure, add the Ethernet interfaces of Switch A and Switch B to corresponding vlan, and start PIM6 on each vlan interface.

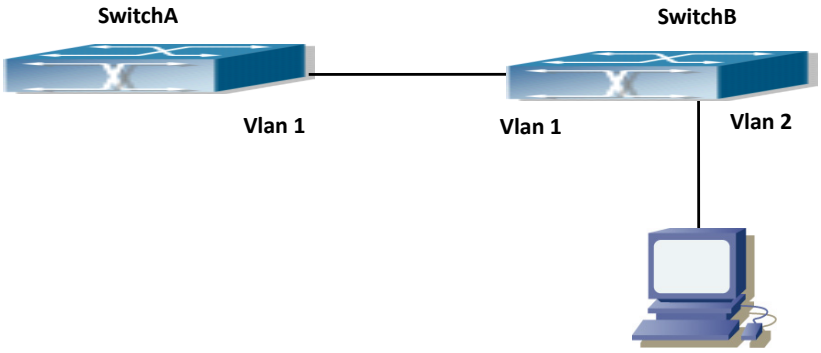


Figure 14-35 Network Topology Diagram

The configuration procedure for SwitchA and SwitchB is as below:

(1) Configure SwitchA:

Switch (config) #ipv6 pim multicast-routing

Switch (config) #ipv6 pim rp-address 3FFE::1

Switch (config) #interface vlan 1

Switch (Config-if-Vlan1) #ipv6 address 3FFE::1/64
Switch (Config-if-Vlan1) #ipv6 pim sparse-mode
(2) Configure SwitchB:
Switch (config) #ipv6 pim multicast-routing
Switch (config) #ipv6 pim rp-address 3FFE::1
Switch (config) #interface vlan1
Switch (Config-if-Vlan1) #ipv6 address 3FFE::2/64
Switch (Config-if-Vlan1) #ipv6 pim sparse-mode
Switch (Config-if-Vlan1) #exit
Switch (config) #interface vlan2
Switch (Config-if-Vlan2) #ipv6 address 3FFA::1/64
Switch (Config-if-Vlan2) #ipv6 pim sparse-mode
Switch (Config-if-Vlan2) #ipv6 mld query-timeout 150

14.8.6.4 MLD Troubleshooting Help

When configuring and using MLD protocol, MLD protocol may fail to work normally due to physical connections, incorrect configuration and so on. So, users shall note the following points:

- Assure the physical connection is correct.
- Assure the protocol of interface and link is UP (use show interface command)
- Assure to start one kind of multicast protocol on the interface
- Assure the time of the timers of each router on the same network segment is consistent; usually we recommend the default setting.
- Unicast route shall be used to carry out RPF examination for multicast protocol. So the correctness of unicast route shall be guaranteed above all.

If all attempts fail to solve the problems on MLD, please use debug commands such as debug ipv6 MLD event/packet, and copy DEBUG information in 3 minutes and send to Technology Service Center.

14.8.7 MLD Snooping

14.8.7.1 Introduction to MLD Snooping

MLD, the Multicast Listener Discovery Protocol, is used to realize multicasting in the IPv6. MLD is used by the network equipments such as routers which supports multicast for multicast listener discovery, also used by listeners looking forward to join certain multicast group informing the router to receive data packets from certain multicast address, all of which are done through MLD message exchange. First the router send an MLD Multicast listener Query message through a multicast address which can address all the listeners (namely ff02::1). Once there is a listener who wishes to join the multicast address, it will send a MLD Multicast listener Report back through the multicast address.

MLD Snooping is namely the MLD listening. The switch restricts the multicast traffic from flooding through MLD Snooping, and forward the multicast traffic to ports associated to multicast devices only. The switch listens to the MLD messages between multicast routers and listeners, and maintains the multicast group forwarding list based on the listening result. The switches forwards multicast packets according to the multicast forwarding list

The switch realizes the MLD Snooping function while supporting MLD v2. This way, the user can acquire IPv6 multicast with the switch.

14.8.7.2 MLD Snooping Configuration Task

1. Enable the MLD Snooping function
2. Configure the MLD Snooping

1. Enable the MLD Snooping function

Command	Explanation
Global Mode	
ipv6 mld snooping **no ipv6 mld snooping**	Enable global MLD Snooping, the "**no ipv6 mld snooping**" command disables the global MLD snooping.

2. Configure MLD Snooping

Command	Explanation
Global Mode	
ipv6 mld snooping vlan *<vlan-id>* **no ipv6 mld snooping vlan *<vlan-id>***	Enable MLD Snooping on specific VLAN. The "no" form of this command disables MLD Snooping on specific VLAN.
ipv6 mld snooping vlan *<vlan-id>* limit {group *<g_limit>* | source *<s_limit>*} **no ipv6 mld snooping vlan *<vlan-id>* limit**	Configure the number of the groups in which the MLD Snooping can join, and the maximum number of sources in each group. The "no" form of this command restores to the default.
ipv6 mld snooping vlan *<vlan-id>* l2-general-querier **no ipv6 mld snooping vlan *<vlan-id>* l2-general-querier**	Set the VLAN level 2 general querier, which is recommended on each segment. The "no" form of this command cancels the level 2 general querier configuration.
ipv6 mld snooping vlan *<vlan-id>* mrouter-port interface *<interface –name>* **no ipv6 mld snooping vlan *<vlan-id>* mrouter-port interface *<interface –name>***	Configure the static mrouter port in specific vlan. The "no" form of this command cancels the mrouter port configuration.
ipv6 mld snooping vlan *<vlan-id>* mrouter-port learnpim6 **no ipv6 mld snooping vlan *<vlan-id>* mrouter-port learnpim6**	Enable the function that the specified VLAN learns mrouter-port (according to pimv6 packets), the no command will disable the function.
ipv6 mld snooping vlan *<vlan-id>* mrpt *<value>* **no ipv6 mld snooping vlan *<vlan-id>* mrpt**	Configure the keep-alive time of the mrouter port. The "no" form of this command restores to the default.

ipv6 mld snooping vlan <*vlan-id*> query-interval <*value*> **no ipv6 mld snooping vlan <*vlan-id*> query-interval**	Configure the query interval. The “no” form of this command restores to the default.
ipv6 mld snooping vlan <*vlan-id*> immediate-leave **no ipv6 mld snooping vlan <*vlan-id*> immediate-leave**	Configure immediate leave multicast group function for the MLD Snooping of specific VLAN. The “no” form of this command cancels the immediate leave configuration.
ipv6 mld snooping vlan <*vlan-id*> query-mrsp <*value*> **no ipv6 mld snooping vlan <*vlan-id*> query-mrsp**	Configure the query maximum response period. The “no” form of this command restores to the default.
ipv6 mld snooping vlan <*vlan-id*> query-robustness <*value*> **no ipv6 mld snooping vlan <*vlan-id*> query-robustness**	Configure the query robustness, the “no” form of this command restores to the default.
ipv6 mld snooping vlan <*vlan-id*> suppression-query-time <*value*> **no ipv6 mld snooping vlan <*vlan-id*> suppression-query-time**	Configure the suppression query time. The “no” form of this command restores to the default
Ipv6 mld snooping vlan <*vlan-id*> static-group <*X:X::X:X*> [source <*X:X::X:X*>] interface [ethernet | port-channel] <*IFNAME*> **no ipv6 mld snooping vlan <*vlan-id*> static-group <*X:X::X:X*> [source <*X:X::X:X*>] interface [ethernet | port-channel] <*IFNAME*>**	Configure static-group on specified port of the VLAN. The no form of the command cancels this configuration.

14.8.7.3 MLD Snooping Examples

Scenario 1: MLD Snooping Function

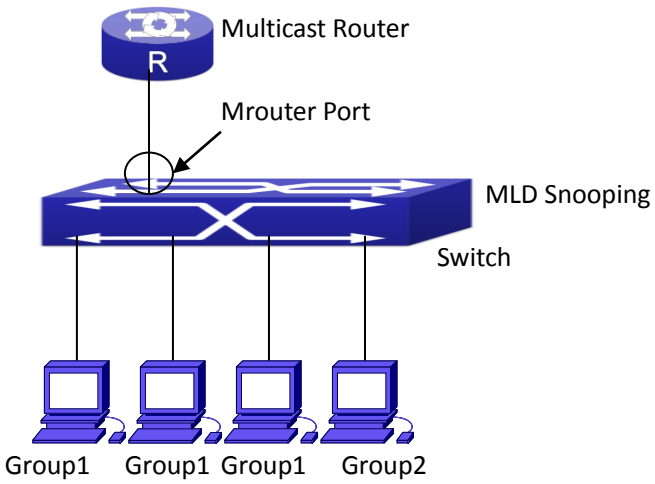


Figure 14-46 Open the switch MLD Snooping Function figure

As shown above, the vlan 100 configured on the switch consists of ports 1, 2, 6, 10 and 12. Four hosts are respectively connected to 2, 6, 10 and 12 while the multicast router on port 1. Suppose we need MLD Snooping on VLAN 100, however by default, the global MLD Snooping as well as the MLD Snooping on each VLAN are, therefore first we have to enable the global MLD Snooping at the same time enable the MLD Snooping on VLAN 100, furthermore we need to set the port 1 of VLAN 100 as a mrouter port.

Configuration procedure is as follows.

Switch#config

Switch(config)#ipv6 mld snooping

Switch(config)#ipv6 mld snooping vlan 100

Switch(config)#ipv6 mld snooping vlan 100 mrouter-port interface ethernet 1/0/1

Multicast configuration:

Assume there are two multicast servers: the Multicast Server 1 and the Multicast Server 2, amongst program 1 and 2 are supplied on the Multicast Server 1 while program 3 on the Multicast server 2, using group addresses respectively the Group 1, Group 2 and Group 3. Concurrently multicast application is operating on the four hosts. Two hosts connected to port 2, 6 are playing program 1 while the host connected to port 10 playing program 2, and the one to port 12 playing program 3.

MLD Snooping interception results:

The multicast table on vlan 100 shows: port 1, 2, 6 are in （Multicasting Server 1, Group1）, port1, 10 are in (Multicasting Server 1,Group2), and port1, 121, 12 are in (Multicasting Server 2, Group3）

All the four hosts successfully receive programs they are interested in. port2, 6 receives no traffic from program2 and 3; port10 receives no traffic from program 1 and 3, and port12 receives no traffic from program1 and 2.

Scenario 2: MLD L2-general-querier

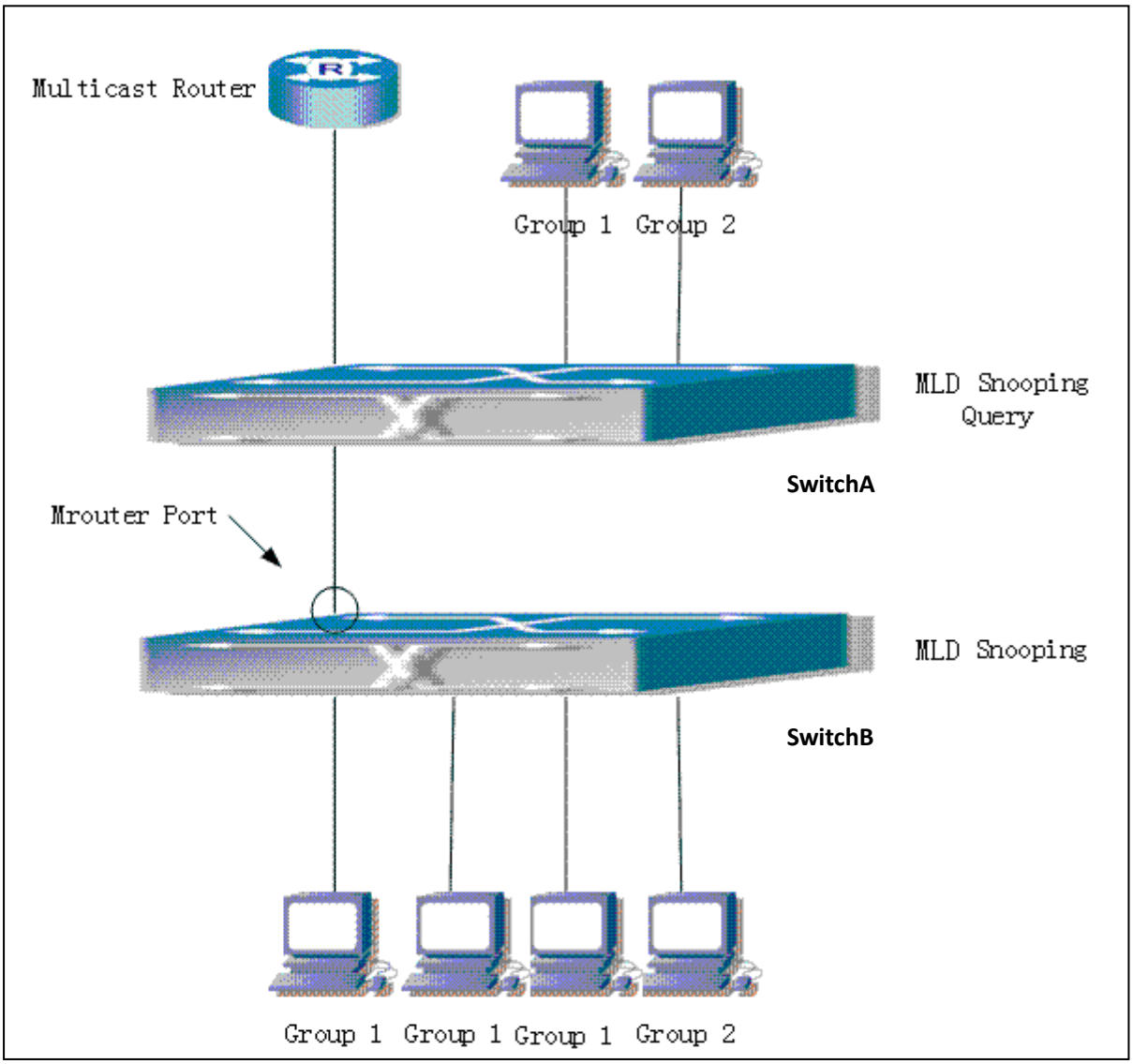


Figure 14-57 Switch as MLD Querier Function figure

Configuration of switch B is the same as the switches in case 1, and here the switch 1 replaces the Multicast Router in case 1. Assume the vlan 60 configured on it contains port 1, 2, 10 and 12, amongst port 1 is connected to multicast server, port 2 to switch2. To send Query periodically, global MLD Snooping has to be enabled while executing the mld snooping vlan 60 l2-general-querier, setting the vlan 60 to a Level 2 General Querier.

Configuration procedure is as follows:

SwitchA#config

SwitchA(config)#ipv6 mld snooping

SwitchA(config)#ipv6 mld snooping vlan 60

SwitchA(config)#ipv6 mld snooping vlan 60 l2-general-querier

SwitchB#config

SwitchB(config)#ipv6 mld snooping

SwitchB(config)#ipv6 mld snooping vlan 100

SwitchB(config)#ipv6 mld snooping vlan 100 mrouter interface ethernet 1/0/1

Multicast configuration:
Same as scenario 1
MLD Snooping interception results:
Same as scenario 1

Scenario 3: To run in cooperation with layer 3 multicast protocols

SWITCH which is used in Scenario 1 is replaced with ROUTER with specific configurations remains the same. And multicast and IGMP snooping configurations are the same with what it is in Scenario 1. To configure PIM-SM6 on ROUTER, and enable PIM-SM6 on vlan 100 (use the same PIM mode with the connected multicast router)

The configurations are listed as below:

```
switch#config
switch(config)#ipv6 pim multicast-routing
switch(config)#interface vlan 100
switch(config-if-vlan100)#ipv6 pim sparse-mode
```

MLD snooping does not distribute entries when layer 3 multicast protocol is enabled. It only does the following tasks.

- To remove the layer 2 multicast entries.
- To provide query functions to the layer 3 with vlan, S, and G as the parameters.
- When layer 3 MLD is disabled, re-enable distributing layer 2 multicast entries.

By looking up the layer 3 IP6MC entries, it can be found that ports can be indicated by the layer 3 multicast entries. This ensures the MLD Snooping can work in cooperation with the layer 3 multicast protocols.

14.8.7.4 MLD Snooping Troubleshooting

In configuring and using MLD Snooping, the MLD Snooping server may fail to run properly due to physical connection failure, wrong configuration, etc. The user should ensure the following:

- ☞ Ensure the physical connection is correct
- ☞ Ensure the MLD Snooping is enabled under global mode (using ipv6 mld snooping)
- ☞ Ensure the MLD Snooping is configured on the vlan under global mode (using ipv6 mld snooping vlan <vlan-id>）
- ☞ Ensure there is a vlan configured as a L2 general querier, or there is a static mrouter configured in a segment,
- ☞ Use command to check if the MLD snooping information is correct

14.9 IPv6 Security RA

14.9.1 Introduction to IPv6 Security RA

In IPv6 networks, the network topology is generally compromised of routers, layer-two switches and IPv6 hosts. Routers usually advertise RA, including link prefix, link MTU and other information, when the IPv6 hosts receive RA, they will create link address, and set the default router as the one sending RA in order to implement IPv6 network communication. If a vicious IPv6 host sends RA to cause that normal IPv6 users set the default router as the vicious IPv6 host user, the vicious user will be able to capture the information of other users, which will threat the network security. Simultaneously, the normal users get incorrect address and will not be able to connect to the network. So, in order to implement the security RA function, configuring on the switch ports to reject vicious RA messages is necessary, thus to prevent forwarding vicious RA to a certain extent and to avoid affecting the normal operation of the network.

14.9.2 IPv6 Security RA Configuration Task Sequence

1. Globally enable IPv6 security RA
2. Enable IPv6 security RA on a port
3. Display and debug the relative information of IPv6 security RA

1. Globally enable IPv6 security RA

Command	Explanation
Global Configuration Mode	
ipv6 security-ra enable **no ipv6 security-ra enable**	Globally enable and disable IPv6 security RA.

2. Enable IPv6 security RA on a port

Command	Explanation
Port Configuration Mode	
ipv6 security-ra enable **no ipv6 security-ra enable**	Enable and disable IPv6 security RA in port configuration mode.

3. Display and debug the relative information of IPv6 security RA

Command	Explanation
Admin Mode	
debug ipv6 security-ra **no debug ipv6 security-ra**	Enable the debug information of IPv6 security RA module, the no operation of this command will disable the output of debug information of IPv6 security RA.
show ipv6 security-ra [interface ***<interface-list>*****]**	Display the distrust port and whether globally security RA is enabled.

14.9.3 IPv6 Security RA Typical Examples

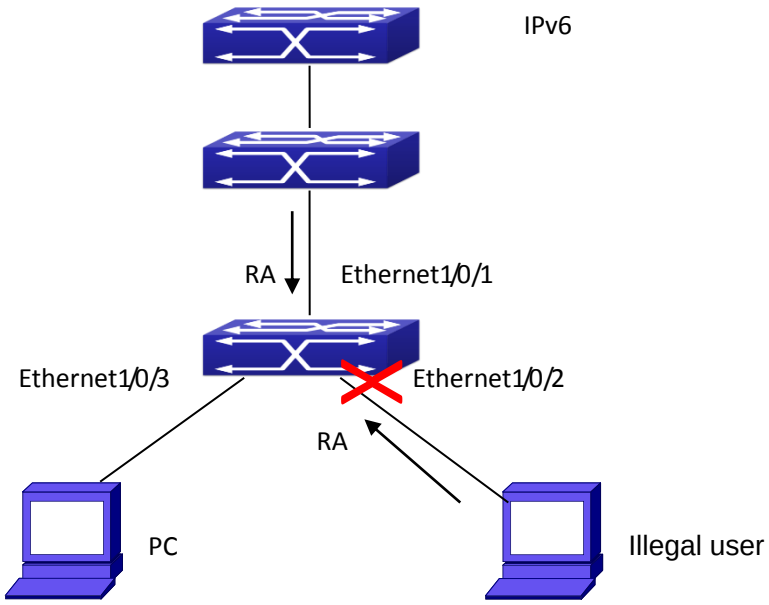


Figure 14-68 IPv6 Security RA sketch map

Instructions: if the illegal user in the graph advertises RA, the normal user will receive the RA, set the default router as the vicious IPv6 host user and change its own address. This will cause the normal user to not be able to connect the network. We want to set security RA on the 1/0/2 port of the switch, so that the RA from the illegal user will not affect the normal user.

Switch configuration task sequence:

Switch#config

Switch(config)#ipv6 security-ra enable

Switch(Config-If-Ethernet1/0/2)# ipv6 security-ra enable

14.9.4 IPv6 Security RA Troubleshooting Help

The function of IPv6 security RA is quite simple, if the function does not meet the expectation after configuring IPv6 security RA:

- Check if the switch is correctly configured.
- Check if there are rules conflicting with security RA function configured on the switch, this kind of rules will cause RA messages to be forwarded.

14.10 SAVI Configuration

14.10.1 Introduction to SAVI

SAVI (Source Address Validation Improvement) is a security authentication method that provides the granularity level of the node source address. It gets the trust node information (such as port, MAC address information), namely, anchor information by monitoring the interaction process of the relative protocol packets (such as ND protocol, DHCPv6 protocol) and using CPS (Control Packet Snooping) mechanism. After that, it binds the anchor information with the node source address and sends the corresponding filter rules, allow the packets which match the filter rules to pass only, so as to reach the aim that check the validity of node source address.

SAVI function includes ND Snooping function, DHCPv6 Snooping function and RA Snooping according to the protocol packet type. ND Snooping function is used to detect ND protocol packet, it sets IPv6 address binding obtained by nodes with the stateless address configuration. DHCPv6 Snooping function is used to detect DHCPv6 protocol packet, it sets IPv6 address binding obtained by nodes with the stateful address configuration. RA Snooping function is used to avoid the lawless node sending the spurious RA packet.

14.10.2 SAVI Configuration

SAVI configuration task list:

1. Enable or disable SAVI function
2. Enable or disable application scene function for SAVI
3. Configure SAVI binding function
4. Configure the global max-dad-delay for SAVI
5. Configure the global max-dad-prepare-delay for SAVI
6. Configure the global max-slaac-life for SAVI
7. Configure the lifetime period for SAVI bind-protect
8. Enable or disable SAVI prefix check function
9. Configure IPv6 address prefix for a link
10. Configure the filter entry number of IPv6 address
11. Configure the check mode for SAVI conflict binding
12. Enable or disable user authentication
13. Enable or disable DHCPv6 trust of port
14. Enable or disable ND trust of port
15. Configure the binding number

1. Enable or disable SAVI function

Command	Explanation
Global mode	

savi enable **no savi enable**	Enable the global SAVI function, no command disables the function.

2. Enable or disable application scene function for SAVI

Command	Explanation
Global mode	
savi ipv6 {dhcp-only | slaac-only | dhcp-slaac} enable **no savi ipv6 {dhcp-only | slaac-only | dhcp-slaac} enable**	Enable the application scene function for SAVI, no command disables the function.

3. Configure SAVI binding function

Command	Explanation
Global mode	
savi ipv6 check source binding ip <*ip-address*> mac <mac-address> interface <*if-name*> {type [slaac | dhcp] lifetime <*lifetime*> | type static} **no savi ipv6 check source binding ip <i*p-address*> interface <*if-name*>**	Configure a static or dynamic binding manually, no command deletes the configured binding. This command may be configured in a global function of savi enable, slaac-only enable, dhcp-only enable or dhcp-slaac enable.

4. Configure the global max-dad-delay for SAVI

Command	Explanation
Global mode	
savi max-dad-delay <*max-dad-delay*> **no savi max-dad-delay**	Configure the max lifetime period of SAVI binding at DETECTION state, no command restores the default value.

5. Configure the global max-dad-prepare-delay for SAVI

Command	Explanation
Global mode	
savi max-dad-prepare-delay <*max-dad-prepare-delay*> **no savi max-dad-prepare-delay**	Configure the max redetection lifetime period for SAVI binding, no command restores the default value.

6. Configure the global max-slaac-life for SAVI

Command	Explanation
Global mode	
savi max-slaac-life <*max-slaac-life*> **no savi max-slaac-life**	Configure the lifetime period of the dynamic slaac binding at BOUND state, no command restores the default value.

7. Configure the lifetime period for SAVI bind-protect

Command	Explanation
Global mode	
savi timeout bind-protect <*protect-time*> **no savi timeout bind-protect**	Configure the bind-protect lifetime period to a port after its state from up to down, no command restores the default value.

8. Enable or disable SAVI prefix check function

Command	Explanation
Global mode	
ipv6 cps prefix check enable **no ipv6 cps prefix check enable**	Enable the address prefix check for SAVI, no command disables the function.

9. Configure IPv6 address prefix for a link

Command	Explanation
Global mode	
ipv6 cps prefix <*ip-address*> vlan <*vid*> **no ipv6 cps prefix <*ip-address*>**	Configure IPv6 address prefix for a link manually, no command deletes the configured address prefix.

10. Configure the filter entry number of IPv6 address

Command	Explanation
Global mode	
savi ipv6 mac-binding-limit <limit-num> **no savi ipv6 mac-binding-limit**	Configure the corresponding dynamic binding number for the same MAC address, no command restores the default value. Note: The binding number only limits the dynamic binding, but does not limit the static binding number.

11. Configure the check mode for SAVI conflict binding

Command	Explanation
Global mode	
**savi check binding <*simple	probe*> mode** **no savi check binding mode**	Configure the check mode for the conflict binding, no command deletes the check mode.

12. Enable or disable user authentication

Command	Explanation
Port mode	
**savi ipv6 check source [ip-address mac-address	ip-address	mac-address]** **no savi ipv6 check source**	Enable the control authentication function for user, no command disables the function.

13. Enable or disable DHCPv6 trust of port

Command	Explanation
Port mode	
ipv6 dhcp snooping trust **no ipv6 dhcp snooping trust**	Enable DHCPv6 trust port, no command disables the trust function. (port is translated from trust port into untrust port)

14. Enable or disable ND trust of port

Command	Explanation
Port mode	
ipv6 nd snooping trust **no ipv6 nd snooping trust**	Configure a port as slaac trust and RA trust, no command deletes the port's trust function.

15. Configure the binding number

Command	Explanation
Port mode	
savi ipv6 binding num <*limit-num*> **no savi ipv6 binding num**	Configure the binding number of a port, no command restores the default value. Note: The binding number only limits the dynamic binding, but does not limit the static binding number.

14.10.3 SAVI Typical Application

In actual application, SAVI function is usually applied in access layer switch to check the validity of node source address on direct-link. There are four typical application scenes for SAVI function: DHCP-Only, Slaac-Only, DHCP-Slaac and Static binding. In network environment, users can select the corresponding scene according to the actual requirement; in double stacks network, while SAVI function associates with IPv4 DHCP snooping to use, IPv4 and IPv6 source address authentication is implemented.

Typical network topology application for SAVI function:

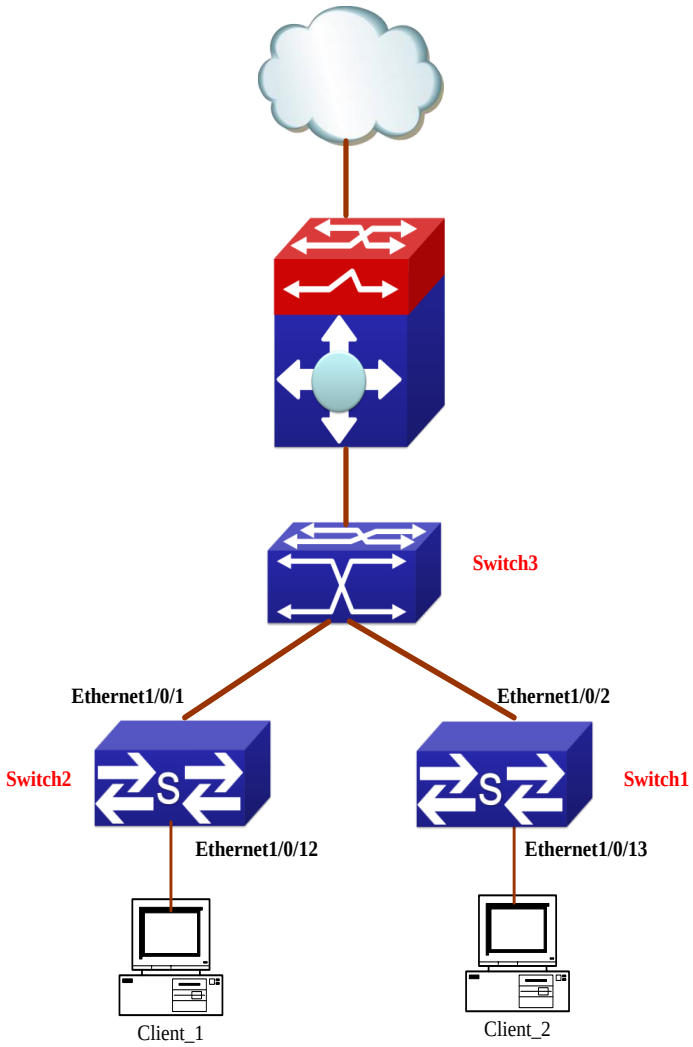


Figure 14-79

Client_1 and Client_2 means two different user's PC installed IPv6 protocol, respectively connect with port Ethernet1/0/12 of Switch1 and port Ethernet1/0/13 of Switch2, and enable the source address check function of SAVI. Ethernet1/0/1 and Ethernet1/0/2 are uplink ports of Switch1 and Switch2 respectively, enable DHCP trust and ND trust functions. Aggregation Switch3 enables DHCPv6 server function and route advertisement function.

Configuration steps of SAVI DHCP-SLAAC scene:

Switch1>enable
Switch1#config
Switch1(config)#savi enable
Switch1(config)#savi ipv6 dhcp-slaac enable
Switch1(config)#savi check binding probe mode
Switch1(config)#interface ethernet1/0/1
Switch1(config-if-ethernet1/0/1)#ipv6 dhcp snooping trust
Switch1(config-if-ethernet1/0/1)#ipv6 nd snooping trust
Switch1(config-if-ethernet1/0/1)#exit
Switch1(config)#interface ethernet1/0/12-20

```
Switch1(config-if-port-range)#savi ipv6 check source ip-address mac-address
Switch1(config-if-port-range)#savi ipv6 binding num 4
Switch1(config-if-port-range)#exit
Switch1(config)#exit
Switch1#write
```

14.10.4 SAVI Troubleshooting

After ensure no problem about SAVI client hardware and cable, please check the status which may exist and the propositional solutions in the following:

- ☞ If IPv6 packets are filtered incorrectly after enable SAVI function, please ensure the global SAVI function enabled. After that, enable the global function of the corresponding SAVI scene according to the actual application scene and enable the port authentication function.
- ☞ If client can not correctly obtain IPv6 address assigned by DHCPv6 server after enable SAVI function, please ensure DHCP port trust is configured by uplink port with DHCPv6 server.
- ☞ If node binding can not be set for the new user after enable SAVI function, please check whether the direct-link port configures the max binding number, and whether the binding number reaches to the max number. If the binding number exceeds the max binding limit, it is recommended to configure the bigger binding limit.
- ☞ If node binding can not be set for new user after configure the bigger binding limit, please check whether the direct-link port configures the corresponding binding number, and whether the corresponding binding number reaches to the max number in the same MAC address. If the binding number exceeds the max binding limit, it is recommended to configure the bigger binding limit.

14.11 IPv6 VRRPv3

14.11.1 Introduction to VRRPv3

VRRPv3 is a virtual router redundancy protocol for IPv6. It is designed based on VRRP (VRRPv2) in IPv4 environment. The following is a brief introduction to it.

In a network based on TCP/IP protocol, in order to guarantee the communication between the devices which are not physically connected, routers should be specified. At present there are two most commonly used methods to specify routers: one is to study dynamically via routing protocols (such as internal routing protocols RIP and OSPF); the other is to configure statically. Running dynamical routing protocol on each terminal is unrealistic, since most operating systems for client end do not support dynamical routing protocol, even if they do, they are limited by the overheads of management, convergence, security and many other problems. So the common method is to adopt static routing configuration on terminal IP devices, which usually means specify one or more default gateway for terminal devices. Static routing simplifies the

management of network and reduces the communication overheads of terminal devices, but it still has a disadvantage: if the router acting as the default gateway breaks, the communication of all the hosts which use this gateway as their next hop host. Even if there are more than one default gateways, before rebooting the terminal devices, they can not switch to the new gateway. Adopting virtual router redundancy protocol (VRPR) can effectively avoid the flaws of statically specifying gateways.

In VRRP protocol, there are two groups of import concepts: VRRP routers and virtual routers, master routers and backup routers. VRRP routers are routers running VRRP, which are physical entities; virtual routers are the ones created by VRRP, which are logical concepts. A group of VRRP routers cooperate to comprise a virtual router, which acts outwardly as a logical router with a unique fixed IP address and MAC address. The routers belonging to the same VRRP group play two mutually exclusive roles at the same time: master routers and backup routers. One VRRP group can only have one master router other but one or more backup routers. VRRPv3 protocol uses selection policy to select a master router from the router group to take charge of responding ND(Neighbor Discovery) neighbor request messages(ARP in IPv4) and forwarding IP data packets, while the other routers in the group will be in a state of waiting as backups. When the master router has a problem for some season, the backup router will be updated to the master router after a delay of a few seconds. Since this switch is very fast and does not need to change IP address or MAC address, it will be transparent to terminal user systems.

In IPv6 environment, the hosts in a LAN usually learn the default gateway via neighbor discovery protocol (NDP), which is implemented based on regularly receiving advertisement messages from routers. The NDP of IPv6 has a mechanism called Neighbor Unreachability Detection, which checks whether a neighbor node is failed by sending unicast neighbor request messages to it. In order to reduce the overheads of sending neighbor request messages, these messages are only sent to those neighbor nodes which are sending flows, and are only sent if there is no instruction of UP state of the router in a period of time. In Neighbor Unreachability Detection, if adopting default parameters, it will take about 38 seconds to detect an unreachable router, which is a delay not ignorable for users and might cause a time-out in some transport protocols. Compared with NDP, VRRP provides a fast default gateway switch. In VRRP, backup routers can take up the unavailable master router in about 3 seconds (default parameter), and this process needs no interaction with hosts, which means being transparent to hosts.

14.11.1.1 The Format of VRRPv3 Message

VRRPv3 has its own message format, VRRP messages are used to communicate the priority of routers and the state of Master in the backup group, they are encapsulated in IPv6 messages to send, and are sent to the specified IPv6 multicast address. The format of VRRPv3 message is shown in Graph 1. The source address of the IPv6 message encapsulating the VRRPv3 message is the local address of the outbound interface of the message, and the destination address of it is the IPv6 multicast address(the multicast allocated to VRRPv3 is FF02:0:0:0:0:0:0:12). The number of hops should be limited to 255, and the next message head is 112(representing a VRRP message).

The meaning of each field in a VRRPv3 message is shown as follows:

Version: The version of VRRPv3, whose value is 3;

Type: The type of VRRP messages. There is only one type: ADVERTISEMENT, and its value is 1;

Virtual Rtr ID：The ID of the virtual router;

Priority：Priority, ranging from 0 to 255;

Count IPv6 Addr：The number of IPv6 addresses in a VRRPv3 message, the minimum of which is 1;

Rsvd：Reserved field, whose value is 0;

Adver Int：The advertisement interval of VRRPv3 messages, in seconds;

Checksum：The checksum, taking account of the whole VRRPv3 message and an IPv6 pseudo head (please refer to RFC2460 for details);

IPv6 Address(es)：one or more IPv6 addresses related to the virtual router, the number of which is the same with ”Count IPv6 Addr”, and the first one of which should be the virtual IPv6 address of the virtual router.

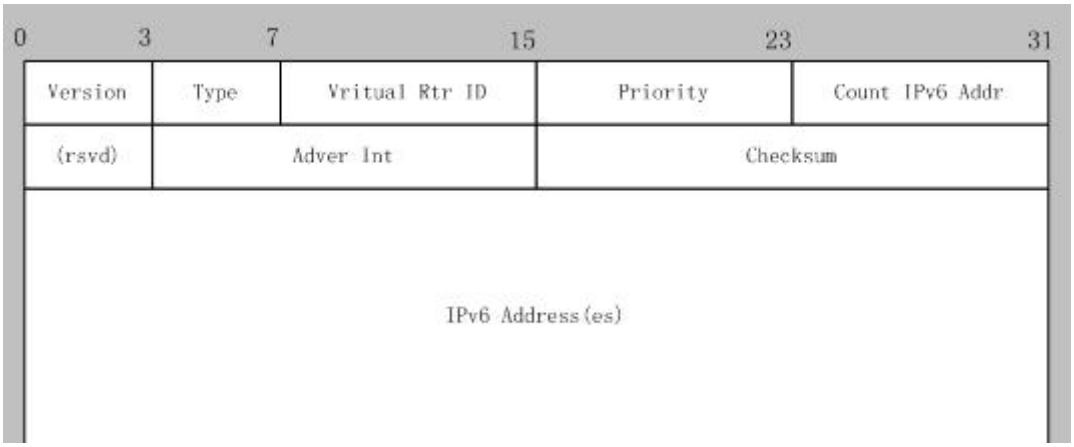


Figure 14-20 VRRPv3 message

14.11.1.2 VRRPv3 Working Mechanism

The working mechanism of VRRPv3 is the same with that of VRRPv2, which is mainly implemented via the interaction of VRRP advertisement messages. It will be briefly described as follows:

Each VRRP router has a unique ID: VRIP, ranging from 1 to 255. This router has a unique virtual MAC address outwardly, and the format of which is 00-00-5E-00-02-{VRID} (the format of virtual MAC address in VRRPv2 is 00-00-5E-00-01-{VRID}). Master router is in charge of using this MAC address to respond to ND neighbor request (it is ARP request in VRRPv2). Thus, no matter what switch is made, the terminal devices will get the same IP and MAC address all the time, reducing the affection that the switch causes on terminal devices.

There is only one kind of VRRP control message: VRRP advertisement. It uses IP multicast data packets to encapsulate, and the format of multicast addresses is FF02:0:0:0:0:0:XXXX:XXXX. In order to keep a consistence with the multicast address in VRRPv2 (224.0.0.18), the multicast addresses used by VRRPv3 advertisement messages can be FF02:0:0:0:0:0:0:12, and the advertisement is limited within the same LAN. Thus, different VRID are guaranteed to be used repeatedly in different networks. In order to reduce the overheads of network bandwidth, only master routers can send VRRP advertisement messages regularly. Backup routers will start a new

round of VRRP selection if it hasn't received a VRRP advertisement in 3 advertisement intervals in a row or if it receives an advertisement with a priority of 0.

In a VRRP router group, the master router is selected according to priority. The range of priority in VRRP protocol is 0-255. If the IP address of a VRRP router is the same to that of the virtual router interface, then the virtual router will be called the IP address owner in the VRRP group; the IP address owner automatically has the highest priority: 255. The priority of 0 is usually used when the IP address owner gives up the role of master. The range of priority can be configured is 1-254. The configuration rule of priority can be set according to the speed and cost of the link, the performance and reliability of the router and other management policies. In the selection of the master router, the virtual router with high priority will win. So, if there is an IP owner in the VRRP group, it will always be the master router. For the candidate routers having the same priority, selection will be done according to the magnitude of IP addresses (the bigger IP address takes precedence). VRRP also provides a preemptive priority policy. If such policy is configured, the backup router with higher priority will preempt the role of new master router over the current master router with lower priority.

In order to avoid the fault of returning a physical MAC address when Pinging virtual IP, it is regulated that virtual IP can not be the real IP of the interface. Thus, all the interfaces participating of the backup group selection will be backup by default.

14.11.2 VRRPv3 Configuration

14.11.2.1 Configuration Task Sequence

1. Create/delete the virtual router (necessary)
2. Configure the virtual IPv6 address and interface of VRRPv3 (necessary)
3. Enable/disable the virtual router (necessary)
4. Configure VRRPv3 assistant parameters (optional)
   (1) Configure VRRPv3 preempt mode
   (2) Configure VRRPv3 priority
   (3) Configure the VRRPv3 advertisement interval
   (4) Configure the monitor interface of VRRPv3

1. Create/delete the virtual router

Command	Explanation
Global Configuration Mode	
router ipv6 vrrp <*vrid*> **no router ipv6 vrrp <*vrid*>**	Create/delete the virtual router.

2. Configure the virtual IPv6 address and interface of VRRPv3

Command	Explanation
VRRPv3 Protocol Mode	

virtual-ipv6 <*ipv6-address*> Interface {Vlan <*ID*> | *IFNAME* } **no virtual-ipv6 interface**	Configure the virtual IPv6 address and interface of VRRPv3, the no operation of this command will delete the virtual IPv6 address and interface.

3. Enable/disable the virtual router

Command	Explanation
VRRPv3 Protocol Mode	
enable	Enable the virtual router.
disable	Disable the virtual router.

4. Configure VRRPv3 assistant parameters

(1) Configure VRRPv3 preempt mode

Command	Explanation
VRRPv3 Protocol Mode	
preempt-mode {true| false}	Configure VRRPv3 preempt mode.

(2) Configure VRRPv3 priority

Command	Explanation
VRRPv3 Protocol Mode	
priority < *priority* >	Configure VRRPv3 priority.

(3) Configure the VRRPv3 advertisement interval

Command	Explanation
VRRPv3 Protocol Mode	
advertisement-interval <time>	Configure the VRRPv3 advertisement interval (in cent seconds).

(4) Configure the monitor interface of VRRPv3

Command	Explanation
VRRPv3 Protocol Mode	
circuit-failover {vlan <*ID*>| *IFNAME*} <*value_reduced*> **no circuit-failover**	Configure the monitor interface of VRRPv3, the no operation of this command will delete the monitor interface.

14.11.3 VRRPv3 Typical Examples

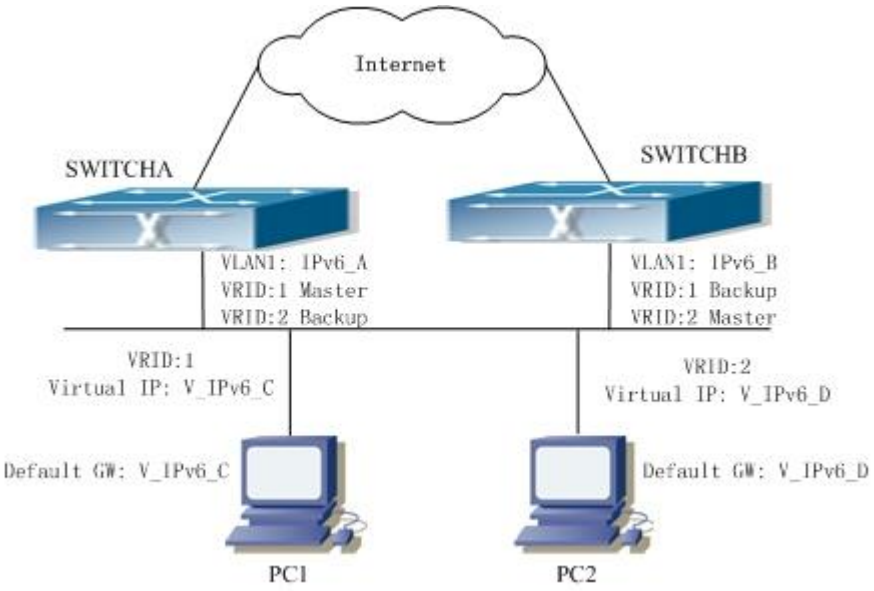


Figure 14-21 VRRPv3 Typical Network Topology

As shown in graph, switch A and switch B are backups to each other, switch A is the master of backup group 1 and a backup of backup group 2. Switch B is the master of backup group 2 and a Backup of backup group 1. The IPv6 addresses of switch A and switch B are “IPv6_A” and “IPv6_B” respectively (it is recommended that IPv6_A and IPv6_B are in the same segment), the virtual IPv6 address of backup group 1 and backup group are “V_IPv6_C” and “V_IPV6_D” respectively, and the default IPv6 gateway address are configured as “V_IPv6_C” and “V_IPv6_D” respectively (in reality, the IPv6 gateway address of hosts are usually learnt automatically via router advertisements, thus, the IPv6 next hop of the hosts will have some randomness). Doing this will not only implement router backup but also the flow sharing function in the LAN.

The configuration of SwitchA:

```
SwitchA (config)#interface vlan 1
SwitchA (config)#router ipv6 vrrp 1
SwitchA (config-router)#virtual-ipv6 fe80::2 interface vlan 1
SwitchA (config-router)#priority 150
SwitchA (config-router)#enable
SwitchA (config)#router ipv6 vrrp 2
SwitchA (config-router)#virtual-ipv6 fe80::3 interface vlan 1
SwitchA (config-router)#enable
```

The configuration of SwitchB:

```
SwitchB (config)# interface vlan 1
SwitchB (config)# router ipv6 vrrp 2
SwitchB (config-router)# virtual-ipv6 fe80::3 interface vlan 1
SwitchB (config-router)# priority 150
SwitchB (config-router)# enable
SwitchB (config)# router ipv6 vrrp 1
SwitchB (config-router)# virtual-ipv6 fe80::2 interface vlan 1
```

SwitchB (config-router)# enable

14.11.4 VRRPv3 Troubleshooting

When configuring and using VRRPv3 protocol, it might operate abnormally because of incorrect physical connections and configuration. So, users should pay attention to the following points:

- First, the physical connections should be correct;
- Next, the interface and link protocol are UP (use **show ipv6 interface** command);
- And then, make sure that IPv6 forwarding function is enabled (use **ipv6 enable** command);
- Besides, make sure that VRRPv3 protocol is enable on the interface;
- Check whether the time of timer in different routers (or layer-three Ethernet switch) within the same backup group is the same;
- Check whether the virtual IPv6 addresses in the same backup group is the same.