NETGEAR®

Software Administration Manual

M4500 Intelligent Fully Managed Switches

Software Version 7.0.0

Model M4500-32C

Model M4500-48XF8C

July 2020
202-12039-02

NETGEAR, Inc.
350 E. Plumeria Drive
San Jose, CA 95134, USA

Support and Community

Visit netgear.com/support to get your questions answered and access the latest downloads.

You can also check out our NETGEAR Community for helpful advice at community.netgear.com.

Regulatory and Legal

Si ce produit est vendu au Canada, vous pouvez accéder à ce document en français canadien à https://www.netgear.com/support/download/.

(If this product is sold in Canada, you can access this document in Canadian French at https://www.netgear.com/support/download/.)

For regulatory compliance information including the EU Declaration of Conformity, visit https://www.netgear.com/about/regulatory/.

See the regulatory compliance document before connecting the power supply.

For NETGEAR's Privacy Policy, visit https://www.netgear.com/about/privacy-policy.

By using this device, you are agreeing to NETGEAR's Terms and Conditions at https://www.netgear.com/about/terms-and-conditions. If you do not agree, return the device to your place of purchase within your return period.

Trademarks

Revision History

Publication Part Number	Publish Date	Comments
202-12039-02	July 2020	We added the PTP End-to-End Transparent Clock feature.
202-12039-01	September 2019	First publication.

Contents

1. Supported Features on the M4500 Series Switches 12

1.1. Switching Features Introduction 12

1.1.1. VLAN Support 12
1.1.2. Double VLANs 12
1.1.3. Switching Modes 12
1.1.4. Spanning Tree Protocols (STP) 12
1.1.5. Rapid Spanning Tree 12
1.1.6. Multiple Spanning Tree 13
1.1.7. Bridge Protocol Data Unit (BPDU) Guard 13
1.1.8. Port-channel 13
1.1.9. Link Aggregate Control Protocol (LACP) 13
1.1.10. Multi Chassis Link Aggregation Group (MLAG) 13
1.1.11. Flow Control Support (IEEE 802.3x) 13
1.1.12. Asymmetric Flow Control 14
1.1.13. Alternate Store and Forward (ASF) 14
1.1.14. Jumbo Frames Support 14
1.1.15. Auto-MDI/MDIX Support 14
1.1.16. Unidirectional Link Detection (UDLD) 14
1.1.17. Expandable Port Configuration 14
1.1.18. VLAN-aware MAC-based Switching 15
1.1.19. Back Pressure Support 15
1.1.20. Auto Negotiation 15
1.1.21. Storm Control 15
1.1.22. Port Mirroring 15
1.1.23. sFlow 16
1.1.24. Static and Dynamic MAC Address Tables 16
1.1.25. Link Layer Discovery Protocol (LLDP) 16
1.1.26. Link Layer Discovery Protocol (LLDP) for Media Endpoint Device 16
1.1.27. DHCP Layer 2 Relay 16
1.1.28. MAC Multicast Support 16
1.1.29. IGMP Snooping 17
1.1.30. SDVoE 17
1.1.31. Source Specific Multicasting (SSM) 17
1.1.32. Control Packet Flooding 17

1.1.33. Flooding to mRouter Ports 17
1.1.34. IGMP Snooping Querier 17
1.1.35. Management and Control Plane ACLs 18
1.1.36. Remote Switched Port Analyzer (RSPAN) 18
1.1.37. Link Dependency 18
1.1.38. IPv6 Router Advertisement Guard 18
1.1.39. FIP Snooping 19
1.1.40. ECN Support 19
1.2. Security Features 20
1.2.1. Configurable Access and Authentication Profiles 20
1.2.2. AAA Command Authorization 20
1.2.3. Password-protected Management Access 20
1.2.4. Strong Password Enforcement 20
1.2.5. MAC-based Port Security 20
1.2.6. RADIUS Client 20
1.2.7. TACACS+ Client 20
1.2.8. Dot1x Authentication (IEEE 802.1X) 21
1.2.9. MAC Authentication Bypass 21
1.2.10. DHCP Snooping 21
1.2.11. DHCPv6 Snooping 21
1.2.12. Dynamic ARP Inspection 22
1.2.13. IP Source Address Guard 22
1.3. Quality of Service Features 22
1.3.1. Access Control Lists (ACL) 22
1.3.2. ACL Remarks 22
1.3.3. ACL Rule Priority 22
1.3.4. Differentiated Service (DIffServ) 23
1.3.5. Class of Service (CoS) 23
1.4. Management Features 23
1.4.1. Management Options 23
1.4.2. Management of Basic Network Information 23
1.4.3. File Management 23
1.4.4. Malicious Code Detection 24
1.4.5. Automatic Installation of Firmware and Configuration 24
1.4.6. Warm Reboot 24
1.4.7. SNMP Alarms and Trap Logs 24

1.4.8. Remote Monitoring (RMON) 24
1.4.9. Statistics Application 24
1.4.10. Log Messages 25
1.4.11. System Time Management 25
1.4.12. Source IP Address Configuration 25
1.4.13. Multiple Linux Routing Tables 25
1.4.14. Open Network Install Environment Support 25
1.4.15. Interface Error Disable and Auto Recovery 25
1.4.16. CLI Scheduler 26
1.5. Routing Features 26
1.5.1. IP Unnumbered 26
1.5.2. Open Shortest Path First (OSPF) 26
1.5.3. Border Gateway Protocol (BGP) 26
1.5.4. VLAN Routing 27
1.5.5. IP Configuration 27
1.5.6. Address Resolution Protocol (ARP) Table Management 28
1.5.7. BOOTP/DHCP Relay Agent 28
1.5.8. IP Helper and UDP Relay 28
1.5.9. Routing Table 28
1.5.10. Virtual Router Redundancy Protocol (VRRP) 28
1.5.11. Algorithmic Longest Prefix Match (ALPM) 28
1.5.12. Bidirectional Forwarding Detection 28
1.5.13. VRF Lite Operation and Configuration 29
1.6. Layer 3 Multicast Features 29
1.6.1. Internet Group Management Protocol 29
1.6.2. Protocol Independent Multicast 29
1.6.3. MLD/MLDv2 (RFC2710/RFC3810) 30
1.7. Data Center Features 30
1.7.1. Priority-Based Flow Control 30
1.7.2. Data Center Bridging Exchange Protocol 30
1.7.3. CoS Queuing and Enhanced Transmission Selection 30
1.7.4. VXLAN Gateway 30
2. Getting Started 32
2.1. Accessing the switch Command-Line Interface 32
2.1.1. Connecting to the Switch Console 32
2.1.2. Login User ID and Password 33

2.1.3. Accessing the Switch CLI through the Network 33
2.1.4. Using the Service Port or Management VLAN Interface for Remote Management 34
2.1.5. DHCP Option 61 35
2.2. Understanding the User Interfaces 36
2.2.1. Using the Command-Line Interface 37
2.2.2. Using SNMP 37
3. Configuring L2 Switching Features 43
3.1. Port Configuration 43
3.1.1. 100G Port-mode Command 43
3.2. Virtual Local Area Networks 44
3.2.1. VLAN Tagging 45
3.2.2. Double-VLAN Tagging 46
3.2.3. Default VLAN Behavior 47
3.2.4. VLAN Configuration Example 47
3.3. Switchport Modes 51
3.4. Port-channels – Operation and Configuration 53
3.4.1. Static and Dynamic Port-channel 53
3.4.2. Port-channel Hashing 54
3.4.3. Port-channel Interface Overview 55
3.4.4. Port-channel Interaction with Other Features 56
3.4.5. Port-channel Configuration Guidelines 57
3.5. LACP Fallback Configuration 60
3.5.1. Configuring Dynamic Port-channels 60
3.5.2. Configuring Static Port-channels 61
3.6. MLAG – Operation and Configuration 62
3.6.1. Overview 62
3.6.2. Deployment Scenarios 63
3.6.3. MLAG Fast Failover 67
3.6.4. MLAG Configuration 67
3.7. Unidirectional Link Detection (UDLD) 70
3.7.1. UDLD Modes 71
3.7.2. UDLD and Port-channel Interfaces 71
3.7.3. Configuring UDLD 71
3.8. Port Mirroring 73
3.8.1. Configuring Port Mirroring 73

3.8.2. Configuring RSPAN 74
3.8.3. VLAN-based Mirroring 76
3.8.4. Flow-based Mirroring 76
3.9. Spanning Tree Protocol 77
3.9.1. Classic STP, Multiple STP, and Rapid STP 77
3.9.2. STP Operation 77
3.9.3. MSTP in the Network 78
3.9.4. Optional STP Features 81
3.9.5. STP Configuring Examples 83
3.10. IGMP Snooping 84
3.10.1. IGMP Snooping Querier 84
3.10.2. Configuring IGMP Snooping 85
3.10.3. IGMPv3/SSM Snooping 88
3.11. SDVoE 88
3.11.1. IGMP & IGMP Snooping Enhancements for IGMP V1 & V2 88
3.11.2. SDVoE Configuration Example 91
3.12. MLD Snooping 93
3.12.1. MLD Snooping Configuration Example 93
3.12.2. MLD Snooping First Leave Configuration Example 96
3.12.3. MLD Snooping Querier Configuration Example 97
3.13. LLDP and LLDP-MED 98
3.13.1. LLDP and Data Center Application 99
3.13.2. Configuring LLDP 99
3.14. sFlow 101
3.14.1. sFlow Sampling 102
3.14.2. Configuring sFlow 103
3.15. Link Dependency 104
3.16. FIP Snooping 105
3.17. ECN 109
3.17.1. Enabling ECN in Microsoft Windows 110
3.17.2. Example 1: SLA Example 110
3.17.3. Example 2: Data Center TCP (DCTCP) Configuration 113
3.18. Storm Control 114
3.18.1. Storm Control Configuration Example 114
3.19. Jumbo Frames 115

3.19.1. Jumbo Frame Configuration Example 115
3.20. Port-Backup 116
3.20.1. Port-Backup Configuration Example 116
3.21. PTP End-to-End Transparent Clock 117
3.21.1. PTP Time Stamp Operation 118
3.21.2. PTP Transparent Clocks 119
3.21.3. Manage the PTP End-to-End Transparent Clock 119
3.21.4. Globally Reenable PTP End-to-End Transparent Clock 120
3.21.5. Reenable PTP End-to-End Transparent Clock for an Interface 120
3.21.6. Display the PTP End-to-End Transparent Clock Status 120
4. Configuring Security Features 122
4.1. Controlling Management Access 122
4.1.1. Using RADIUS Servers for Management Security 122
4.1.2. Using TACACS+ to Control Management Access 123
4.1.3. Configuring and Applying Authentication Profiles 124
4.1.4. Configuring the Primary and Secondary RADIUS Servers 126
4.1.5. Configuring an Authentication Profile 126
4.2. Configuring DHCP Snooping, DAI, and IPSG 128
4.2.1. DHCP Snooping Overview 128
4.2.2. IP Source Guard Overview 130
4.2.3. Dynamic ARP Inspection Overview 131
4.2.4. Increasing Security with DHCP Snooping, DAI, and IPSG 131
4.2.5. Configuring DHCP Snooping 132
4.2.6. Configuring IPSG 133
4.3. Configuring DHCPv6 Snooping 134
4.3.1. DHCPv6 Snooping Configuration Example 134
4.4. ACLs 136
4.4.1. MAC ACLs 137
4.4.2. IP ACLs 137
4.4.3. ACL Redirect Function 138
4.4.4. ACL Mirror Function 138
4.4.5. ACL Logging 138
4.4.6. Time-based ACLs 138
4.4.7. ACL Rule Remarks 139
4.4.8. ACL Rule Priority 139
4.4.9. ACL Limitations 140

4.4.10. ACL Configuration Process 140
4.4.11. Preventing False ACL Matches 140
4.4.12. IPv6 ACL Qualifies 141
4.4.13. ACL Configuration Examples 142
4.5. Control Plane Policing (CoPP) 146
4.5.1. CoPP Configuration Examples 146
5. Configuring Quality of Service 149
5.1. CoS 149
5.1.1. Trusted and Untrusted Port Modes 149
5.1.2. Traffic Shaping on Egress Traffic 149
5.1.3. Defining Traffic Queues 150
5.2. DiffServ 152
5.2.1. DiffServ Functionality and Switch Roles 153
5.2.2. Elements of DiffServ Configuration 153
5.2.3. Configuration DiffServ to Provide Subnets Equal Access to External Network 154
6. Configuring Switch Management Features 156
6.1. Managing Images and Files 156
6.1.1. Supported File Management Methods 157
6.1.2. Uploading and Downloading Files 157
6.1.3. Managing Configuration Files 157
6.1.4. Saving the Running Configuration 159
6.1.5. File and Image Management Configuration Examples 159
6.2. Enabling Automatic System Configuration 163
6.2.1. DHCP Auto Install Process 163
6.2.2. Monitoring and Completing the DHCP Auto Install Process 164
6.2.3. DHCP Auto Install Dependencies 165
6.2.4. Default Auto Install Values 165
6.2.5. Enabling DHCP Auto Install 165
6.3. Configuring System Log Example 166
6.3.1. Example 1 to Add Syslog Host 166
6.3.2. Example 2 to Verify Syslog Host Configuration 166
6.4. Configuring CLI Scheduler (Kron) 169
6.4.1. CLI Scheduler Policy Lists 169
6.4.2. CLI Scheduler Occurrences 170
6.4.3. Configuration Example 170

7. Configuring Routing 171

7.1. Basic Routing and Features 171

7.1.1. VLAN Routing 171

7.1.2. IP Routing Configuration Example 172

7.1.3. IP Unnumbered Configuration Example 175

7.2. OSPF 177

7.2.1. Configuring an OSPF Border Router and Setting Interface Costs 178

7.3. VRRP 180

7.3.1. VRRP Operation in the Network 180

7.3.2. VRRP Configuration Example 182

7.4. IP Helper 187

7.4.1. Relay Agent Configuration Example 189

7.5. Border Gateway Patrol (BGP) 191

7.5.1. BGP Topology 192

7.5.2. BGP Behavior 193

7.5.3. BGP Configuration Example 194

7.6. IPv6 Routing 199

7.6.1. How Does IPv6 Compare with IPv6 199

7.6.2. How are IPv6 Interface Configured 200

7.6.3. Default IPv6 Routing Values 200

7.6.4. Configuring IPv6 Routing Features 201

7.7. ECMP Hash Selection 205

7.8. Bidirectional Forwarding Detection 206

7.8.1. Configuring BFD 206

7.9. VRF Lite Operation and Configuration 207

7.9.1. Route Leaking 208

7.9.2. Adding Leaked Routes 208

7.9.3. Using Leaked Routes 208

7.9.4. CPU-Originated Traffic 208

7.9.5. VRF Features Support 209

7.9.6. VRF Lite Development Scenarios 211

7.9.7. VRF Configuration Example 213

8. Configuring Multicast Routing 215

8.1. L3 Multicast Overview 215

8.1.1. IP Multicast Traffic 215

8.1.2. Multicast Protocol Switch Support 215
8.1.3. Multicast Protocol Roles 216
8.1.4. Multicast Switch Requirements 216
8.1.5. Determining which Multicast Protocols to Enable 216
8.1.6. Multicast Routing Tables 216
8.1.7. Multicast Tunneling 216
8.1.8. IGMP 217
8.1.9. MLD Protocol 217
8.1.10. PIM Protocol 218
8.2. Default L3 Multicast Values 219
8.3. L3 Multicast Configuration Examples 221
8.3.1. Configuring Multicast VLAN Routing with IGMP and PIM-SM 221
8.3.2. Example 1: MLDv1 Configuration 223
8.3.3. Example 2: MLDv2 Configuration 224
8.3.4. Example 3: MLD Configuration Verification 225
9. Configuring Data Center Features 226
9.1. Data Center Technology Overview 226
9.2. Priority-based Flow Control 226
9.2.1. PFC Operation and Behavior 227
9.2.2. Configuring PFC 227
9.3. Data Center Bridging Exchange Protocol 228
9.3.1. Interoperability with IEEE DCBX 229
9.3.2. DCBX and Port Roles 229
9.3.3. Configuration Source Port Selection Process 230
9.3.4. Configuring DCBX 231
9.4. CoS Queuing 232
9.4.1. CoS Queuing Function and Behavior 233
9.4.2. Configuring CoS Queuing and ETS 235
9.5. Enhanced Transmission Selection 237
9.5.1. ETS Operation and Dependencies 237
9.6. VXLAN Gateway Operation and Configuration 238
9.6.1. Overview 238
9.6.2. Functional Description 239
9.6.3. VXLAN Configuration Examples 244
Appendix A: Term and Acronyms 249

1. Supported Features on the M4500 Series Switches

This section provides a brief overview of the supported features on the M4500 Series Switches. The features are categorized as follows:

1.1. Switching Features Introduction

1.1.1. VLAN Support

VLANs are collections of switching ports that comprise a single broadcast domain. Packets are classified as belonging to a VLAN based on either the VLAN tag or a combination of the ingress port and packet contents. Packets sharing common attributes can be groups in the same VLAN. The switch software is in full compliance with IEEE 802.1Q VLAN tagging.

1.1.2. Double VLANs

The Double VLAN feature (IEEE 802.1QinQ) allows the use of a second tag on network traffic. The additional tag helps differentiate between customers in the Metropolitan Area Networks (MAN) while preserving individual customer's VLAN identification when they enter their own 802.1Q domain.

1.1.3. Switching Modes

The switchport mode feature helps to minimize the potential for configuration errors. The feature also makes VLAN configuration easier by reducing the amount of commands needed for port configuration. For example, to configure a port connected to an end user, you can configure the port in Access mode. Ports connected to other switches can be configured in Trunk mode. VLAN assignments and tagging behavior are automatically configured as appropriate for the connection type.

1.1.4. Spanning Tree Protocols (STP)

Spanning Tree Protocol (IEEE 802.1D) is a standard requirement of Layer 2 switches that allows bridges to automatically prevent and resolve L2 forwarding loops. The STP feature supports a variety of per-port settings including path cost, priority settings, Port Fast mode, STP Root Guard, Loop Guard, TCN Guard, and Auto Edge. These settings are also configurable per-Port-channel.

1.1.5. Rapid Spanning Tree

Rapid Spanning Tree Protocol (RSTP) detects and uses network topologies to enable faster spanning tree convergence after a topology change, without creating forwarding loops. The port settings supported by STP are also supported by RSTP.

1.1.6. Multiple Spanning Tree

Multiple Spanning Tree (MSTP) operation maps VLANs to spanning tree instances. Packets assigned to various VLANs are transmitted along different paths within MSTP Regions (MST Regions). Regions are one or more interconnected MSTP bridges with identical MSTP settings. The MSTP standard lets administrators assign VLAN traffic to unique paths.

The switch supports IEEE 802.1Q-2005, which is a version of corrects problems associated with the previous version, provides for faster transition-to-forwarding, and incorporates new features for a port (restricted role and restricted TCN).

1.1.7. Bridge Protocol Data Unit (BPDU) Guard

Spanning Tree BPDU Guard is used to disable the port in case a new device tries to enter the already existing topology of STP. Thus devices, which were originally not a part of STP, are not allowed to influence the STP topology.

1.1.8. Port-channel

Up to 32 ports can combine to form a single Port-Channel. This enables fault tolerance protection from physical link disruption, higher bandwidth connections and improved bandwidth granularity.

A Port-channel is composed of ports of the same speed, set to full-duplex operation.

1.1.9. Link Aggregate Control Protocol (LACP)

Link Aggregate Control Protocol (LACP) uses peer exchanges across links to determine, on an ongoing basis, the aggregation capability of various links, and continuously provides the maximum level of aggregation capability achievable between a given pair of systems. LACP automatically determines, configures, binds, and monitors the binding of ports to aggregators within the system.

1.1.10. Multi Chassis Link Aggregation Group (MLAG)

This feature enables a Port-channel to be created across two independent units, which creates a scenario where some member ports of the MLAG can reside on one unit and the other members of the MLAG can reside on the other unit. The partner device on the remote side can be a MLAG unaware unit. For the MLAG unaware unit, the MLAG appears to be a single Port-channel connected to a single unit.

1.1.11. Flow Control Support (IEEE 802.3x)

Flow control enables lower speed switches to communicate with higher speed switches by requesting that the higher speed switch refrains from sending packets. Transmissions are temporarily halted to prevent buffer overflows.

1.1.12. Asymmetric Flow Control

When in asymmetric flow control mode, the switch responds to PAUSE frames received from peers by stopping packet transmission, but the switch does not initiate MAC control PAUSE frames.

When the switch is configured in asymmetric flow control (or no flow control mode), the device is placed in egress drop mode. Egress drop mode maximizes the throughput of the system at the expense of packet loss in a heavily congested system, and this mode avoids head of line blocking.

Asymmetric flow control is not supported on Fast Ethernet platforms because support was introduced to the physical layer with the Gigabit PHY specifications.

1.1.13. Alternate Store and Forward (ASF)

The Alternate Store and Forward (ASF) feature, which is also known as cut-through mode, reduces latency for large packets. When ASF is enabled, the memory management unit (MMU) can forward a packet to the egress port before it has been entirely received on the Cell Buffer Pool (CBP) memory.

1.1.14. Jumbo Frames Support

Jumbo frames enable transporting data in fewer frames to ensure less overhead, lower processing time, and fewer interrupts. The maximum transmission unit (MTU) size is configurable per-port.

1.1.15. Auto-MDI/MDIX Support

Your switch supports auto-detection between crossed and straight-through cables. Media-Dependent Interface (MDI) is the standard wiring for end stations, and the standard wiring for hubs and switches is known as Media- Dependent Interface with Crossover (MDIX).

1.1.16. Unidirectional Link Detection (UDLD)

The UDLD feature detects unidirectional links physical ports by exchanging packets containing information about neighboring devices. The purpose of the UDLD feature is to detect and avoid unidirectional links. A unidirectional link is a forwarding anomaly in a Layer 2 communication channel in which a bidirectional link stops passing traffic in one direction.

1.1.17. Expandable Port Configuration

Expandable ports allow you to configure a 100GbE port in either 4×25/10GbE mode or 1×40GbE mode. When the 100GbE port is operating in 4×25/10GbE mode, the port operates as four 25/10GbE ports, each on a separate lane. This mode requires the use of a suitable 4×25GbE to 1×100GbE pigtail cable.

Expandable port capability can be enabled on 100G ports using the CLI command **[no] port-mode**. A change to the port mode is made effective immediately.

1.1.18. VLAN-aware MAC-based Switching

Packets arriving from an unknown source address are sent to the CPU and added to the Hardware Table. Future packets addressed to or from this address are more efficiently forwarded.

1.1.19. Back Pressure Support

On half-duplex links, a receiver may prevent buffer overflows by jamming the link so that it is unavailable for additional traffic. On full duplex links, a receiver may send a PAUSE frame indicating that the transmitter should cease transmission of frames for a specified period.

When flow control is enabled, the switch will observe received PAUSE frames or jamming signals, and will issue them when congested.

1.1.20. Auto Negotiation

Auto negotiation allows the switch to advertise modes of operation. The auto negotiation function provides the means to exchange information between two switches that share a point-to-point link segment, and to automatically configure both switches to take maximum advantage of their transmission capabilities.

The switch enhances auto negotiation by providing configuration of port advertisement. Port advertisement allows the system administrator to configure the port speeds that are advertised.

1.1.21. Storm Control

When Layer 2 frames are forwarded, broadcast, unknown unicast, and multicast frames are flooded to all ports on the relevant virtual local area network (VLAN). The flooding occupies bandwidth, and loads all nodes connected on all ports. Storm control limits the amount of broadcast, unknown unicast, and multicast frames accepted and forwarded by the switch.

Per-port and per-storm control type (broadcast, multicast, or unicast), the storm control feature can be configured to automatically shut down a port when a storm condition is detected on the port; or to send a trap to the system log. When configured to shut down, the port is put into a diagnostic-disabled state. The user must manually re-enable the interface for it to be operational. When configured to send a trap, the trap is sent once in every 30 seconds. When neither action is configured, the switch rate-limits the traffic when storm conditions occur.

1.1.22. Port Mirroring

Port mirroring monitors and mirrors network traffic by forwarding copies of incoming and outgoing packets from up to four source ports to a monitoring port. The switch also supports flow-based mirroring, which allows you to copy certain types of traffic to a single destination port. This provides flexibility—instead of mirroring all ingress or egress traffic on a port the switch can mirror a subset of that traffic. You can configure the switch to mirror flows based on certain kinds of Layer 2, Layer 3, and Layer 4 information.

The switch supports up to four monitor sessions. Port mirroring, flow based mirroring, RSPAN, and VLAN mirroring can be configured at the same time on the switch using different sessions IDs and in any

combinations. Any two sessions cannot be identical. Multiple mirroring sessions are supported for all types of mirroring.

A given interface can be used as a source interface for different sessions. For example a mirroring session can be created with source interface as port A and destination interface as port B. Another session can be created with source interface as port A and destination interface as port C. An interface cannot be configured as a destination interface for more than one session.

An IP/MAC access-list can be attached to any mirroring session or to all sessions at the same time.

1.1.23. sFlow

sFlow is the standard for monitoring high-speed switched and routed networks. sFlow technology is built into network equipment and gives complete visibility into network activity, enabling effective management and control of network resources. The switch supports sFlow version 5.

1.1.24. Static and Dynamic MAC Address Tables

You can add static entries to the switch's MAC address table and configure the aging time for entries in the dynamic MAC address table. You can also search for entries in the dynamic table based on several different criteria.

1.1.25. Link Layer Discovery Protocol (LLDP)

The IEEE 802.1AB defined standard, Link Layer Discovery Protocol (LLDP), allows the switch to advertise major capabilities and physical descriptions. This information can help you identify system topology and detect bad configurations on the LAN.

1.1.26. Link Layer Discovery Protocol (LLDP) for Media Endpoint Device

The Link Layer Discovery Protocol for Media Endpoint Devices (LLDP-MED) provides an extension to the LLDP standard for network configuration and policy, device location, Power over Ethernet management, and inventory management.

1.1.27. DHCP Layer 2 Relay

This feature permits Layer 3 Relay agent functionality in Layer 2 switched networks. The switch supports L2 DHCP relay configuration on individual ports, Port-channels and VLANs.

1.1.28. MAC Multicast Support

Multicast service is a limited broadcast service that allows one-to-many and many-to-many connections. In Layer 2 multicast services, a single frame addressed to a specific multicast address is received, and copies of the frame to be transmitted on each relevant port are created.

1.1.29. IGMP Snooping

Internet Group Management Protocol (IGMP) Snooping is a feature that allows a switch to forward multicast traffic intelligently on the switch. Multicast IP traffic is traffic that is destined to a host group. Host groups are identified by class D IP addresses, which range from 224.0.0.0 to 239.255.255.255. Based on the IGMP query and report messages, the switch forwards traffic only to the ports that request the multicast traffic. This prevents the switch from broadcasting the traffic to all ports and possibly affecting network performance.

1.1.30. SDVoE

SDVoE (Software Defined Video-over-Ethernet) is the latest high-performance, software-based AV-over-IP platform for control and distribution of audio and video over Ethernet and fiber networks.

1.1.31. Source Specific Multicasting (SSM)

This mechanism provides the ability for a host to report interest in receiving a particular multicast stream only from among a set of specific source addresses, or its interest in receiving a multicast stream from any source other than a set of specific source addresses.

1.1.32. Control Packet Flooding

This feature enhances the MGMD Snooping functionality to flood multicast packets with DIP=224.0.0.x to all members of the incoming VLAN irrespective of the configured filtering behavior. This enhancement depends on the ability of the switch to flood packets with DIP=224.0.0.x irrespective of the entries in the L2 Multicast Forwarding Tables.

1.1.33. Flooding to mRouter Ports

This feature enhances the MGMD Snooping functionality to flood unregistered multicast streams to all mRouter ports in the VLAN irrespective of the configured filtering behavior. This enhancement depends on the ability of the switch to flood packets to specific ports in the incoming VLAN when there are no entries in the L2 Multicast Forwarding Tables for the specific stream. In platforms that do not have the hardware capability, incoming multicast streams are always flooded in the ingress VLAN when the switch supports an “L2 multicast miss.”

1.1.34. IGMP Snooping Querier

When Protocol Independent Multicast (PIM) and IGMP are enabled in a network with IP multicast routing, the IP multicast router acts as the IGMP querier. However, if it is desirable to keep the multicast network Layer 2 switched only, the IGMP Snooping Querier can perform the query functions of a Layer 3 multicast router.

1.1.35. Management and Control Plane ACLs

This feature provides hardware-based filtering of traffic to the CPU. An optional 'management' feature is available to apply the ACL on the CPU port. Currently, control packets like BPDU are dropped because of the implicit 'deny all' rule added at the end of the list. To overcome this rule, you must add rules that allow the control packets.

Support for user-defined simple rate limiting rule attributes for inbound as well as outbound traffic is also available. This attribute is supported on all QoS capable interfaces - physical, Port-channel, and control-plane.

1.1.36. Remote Switched Port Analyzer (RSPAN)

Along with the physical source ports, the network traffic received/transmitted on a VLAN can be monitored. A port mirroring session is operationally active if and only if both a destination (probe) port and at least one source port or VLAN is configured. If neither is true, the session is inactive. The switch supports remote port mirroring. The switch also supports VLAN mirroring. Traffic from/to all the physical ports which are members of that particular VLAN is mirrored.

Note: The source for a port mirroring session can be either physical ports or VLAN.

For Flow-based mirroring, ACLs are attached to the mirroring session. The network traffic that matches the ACL is only sent to the destination port. This feature is supported for remote monitoring also. IP/MAC access-list can be attached to the mirroring session.

Note: Flow-based mirroring is supported only if QoS feature exists in the package.

Up to four RSPAN sessions can be configured on the switch and up to four RSPAN VLANs are supported. An RSPAN VLAN cannot be configured as a source for more than one session at the same time. To configure four RSPAN mirroring sessions, it is required to configure 4 RSPAN VLANs.

1.1.37. Link Dependency

The Link Dependency feature supports enabling/disabling ports based on the link state of other ports (i.e., making the link state of some ports dependent on the link state of others). In the simplest form, if port A is dependent on port B and switch detects link loss on B, the switch automatically brings down link on port A. When the link is restored to port B, the switch automatically restores link to port A. The link action command option determines whether link A will come up/go down, depending upon the state of link B.

1.1.38. IPv6 Router Advertisement Guard

The switch support IPv6 Router Advertisement Guard (RA-Guard) to protect against attacks via rogue Router Advertisements in accordance with RFC 6105. RA Guard supports Stateless RA-Guard, for which you can configure the interface to allow received router advertisements and router redirect message to be processed/forwarded or dropped.

By default, RA-Guard is not enabled on any interfaces. RA-Guard is enabled/disabled on physical interfaces or Port-channels. RA-Guard does not require IPv6 routing to be enabled.

1.1.39. FIP Snooping

The FCoE Initialization Protocol (FIP) is used to perform the functions of FC_BB_E device discovery, initialization, and maintenance. FIP uses a separate EtherType from FCoE to distinguish discovery, initialization, and maintenance traffic from other FCoE traffic. FIP frames are standard Ethernet size (1518 Byte 802.1q frame), whereas FCoE frames are a maximum of 2240 bytes.

FIP snooping is a frame inspection method used by FIP Snooping Bridges to monitor FIP frames and apply policies based upon the L2 header information in those frames.

FIP snooping allows for:

- Auto-configuration of Ethernet ACLs based on information in the Ethernet headers of FIP frames.
- Emulation of FC point-to-point links within the DCB Ethernet network.
- Enhanced FCoE security/robustness by preventing FCoE MAC spoofing.
- The role of FIP snooping-enabled ports on the switch falls under one of the following types:
    - Perimeter or Edge port (connected directly to a Fiber Channel end node or ENode).
    - Fiber Channel forwarder (FCF) facing port (that receives traffic from FCFs targeted to the ENodes).

Note: The FIP Snooping Bridge feature supports the configuration of the perimeter port role and FCF-facing port roles and is intended for use only at the edge of the switched network.

The default port role in an FCoE-enabled VLAN is as a perimeter port. FCF-facing ports are configured by the user.

1.1.40. ECN Support

Explicit Congestion Notification (ECN) is defined in RFC 3168. Conventional TCP networks signal congestion by dropping packets. A Random Early Discard scheme provides earlier notification than tail drop by dropping packets already queued for transmission. ECN marks congested packets that would otherwise have been dropped and expects an ECN capable receiver to signal congestion back to the transmitter without the need to retransmit the packet that would have been dropped. For TCP, this means that the TCP receiver signals a reduced window size to the transmitter but does not request retransmission of the CE marked packet.

The switch implements ECN capability as part of the WRED configuration process. It is configured as parameter in the **random-detect** command. Eligible packets are marked by hardware based upon the WRED configuration. You can configure any CoS queue to operate in ECN marking mode and can configure different discard thresholds for each color.

1.2. Security Features

1.2.1. Configurable Access and Authentication Profiles

You can configure rules to limit access to the switch management interface based on criteria such as access type and source IP address of the management host. You can also require the user to be authenticated locally or by an external server, such as a RADIUS server.

1.2.2. AAA Command Authorization

This feature enables AAA Command Authorization on the switch.

1.2.3. Password-protected Management Access

Access to the CLI and SNMP management interfaces is password protected, and there are no default users on the system.

1.2.4. Strong Password Enforcement

The Strong Password feature enforces a baseline password strength for all locally administered users. Password strength is a measure of the effectiveness of a password in resisting guessing and brute-force attacks. The strength of a password is a function of length, complexity and randomness. Using strong passwords lowers overall risk of a security breach.

1.2.5. MAC-based Port Security

The port security feature limits access on a port to users with specific MAC addresses. These addresses are manually defined or learned on that port. When a frame is seen on a locked port, and the frame source MAC address is not tied to that port, the protection mechanism is invoked.

1.2.6. RADIUS Client

The switch has a Remote Authentication Dial In User Service (RADIUS) client and can support up to 32 authentication and accounting RADIUS servers.

1.2.7. TACACS+ Client

The switch has a TACACS+ client. TACACS+ provides centralized security for validation of users accessing the switch. TACACS+ provides a centralized user management system while still retaining consistency with RADIUS and other authentication processes.

1.2.8. Dot1x Authentication (IEEE 802.1X)

Dot1x authentication enables the authentication of system users through a local internal server or an external server. Only authenticated and approved system users can transmit and receive data. Supplicants are authenticated using the Extensible Authentication Protocol (EAP). Also supported are PEAP, EAP-TTL, EAP-TTLS, and EAP-TLS.

The switch supports RADIUS-based assignment (via 802.1X) of VLANs, including guest and unauthenticated VLANs. The Dot1X feature also supports RADIUS-based assignment of filter IDs as well as MAC-based authentication, which allows multiple supplicants connected to the same port to each authenticate individually.

1.2.9. MAC Authentication Bypass

The switch supports the MAC-based Authentication Bypass (MAB) feature, which provides 802.1x- unaware clients (such as printers and fax machines) controlled access to the network using the devices' MAC address as an identifier. This requires that the known and allowable MAC address and corresponding access rights be pre-populated in the authentication server. MAB works only when the port control mode of the port is MAC-based.

1.2.10. DHCP Snooping

DHCP Snooping is a security feature that monitors DHCP messages between a DHCP client and DHCP server. It filters harmful DHCP messages and builds a bindings database of (MAC address, IP address, VLAN ID, port) tuples that are specified as authorized. DHCP snooping can be enabled globally and on specific VLANs. Ports within the VLAN can be configured to be trusted or untrusted. DHCP servers must be reached through trusted ports. This feature is supported for both IPv4 and IPv6 packets.

1.2.11. DHCPv6 Snooping

In an IPv6 domain, a node can obtain an IPv6 address using the following mechanisms:

- IPv6 address auto-configuration using router advertisements
- The DHCPv6 protocol

In a typical man-in-the-middle (MiM) attack, the attacker can snoop or spoof the traffic act as a rogue DHCPv6 server. To prevent such attacks, DHCPv6 snooping helps to secure the IPv6 address configuration in the network.

DHCPv6 snooping enables the Brocade device to filter untrusted DHCPv6 packets in a subnet on an IPv6 network. DHCPv6 snooping can ward off MiM attacks, such as a malicious user posing as a DHCPv6 server sending false DHCPv6 server reply packets with the intention of misdirecting other users. DHCPv6 snooping can also stop unauthorized DHCPv6 servers and prevent errors due to user misconfiguration of DHCPv6 servers.

1.2.12. Dynamic ARP Inspection

Dynamic ARP Inspection (DAI) is a security feature that rejects invalid and malicious ARP packets. The feature prevents a class of man-in-the-middle attacks, where an unfriendly station intercepts traffic for other stations by poisoning the ARP caches of its unsuspecting neighbors. The malicious station sends ARP requests or responses mapping another station's IP address to its own MAC address.

1.2.13. IP Source Address Guard

IP Source Guard and Dynamic ARP Inspection use the DHCP snooping bindings database. When IP Source Guard is enabled, the switch drops incoming packets that do not match a binding in the bindings database. IP Source Guard can be configured to enforce just the source IP address or both the source IP address and source MAC address. Dynamic ARP Inspection uses the bindings database to validate ARP packets. This feature is supported for both IPv4 and IPv6 packets.

1.3. Quality of Service Features

1.3.1. Access Control Lists (ACL)

Access Control Lists (ACLs) ensure that only authorized users have access to specific resources while blocking off any unwarranted attempts to reach network resources. ACLs are used to provide traffic flow control, restrict contents of routing updates, decide which types of traffic are forwarded or blocked, and above all provide security for the network. The switch supports the following ALC types:

- IPv4 ACLs
- IPv6 ACLs
- MAC ACLs

For all ACL types, you can apply the ACL rule when the packet enters or exits the physical port, Port-channel, or VLAN interface.

1.3.2. ACL Remarks

Users can use ACL remarks to include comments for ACL rule entries in any MAC ACL. Remarks assist the user in understanding ACL rules easily.

1.3.3. ACL Rule Priority

This feature allows user to add sequence numbers to ACL rule entries and re-sequence them. When a new ACL rule entry is added, the sequence number can be specified so that the new ACL rule entry is placed in the desired position in the access list.

1.3.4. Differentiated Service (DIffServ)

The QoS Differentiated Services (DiffServ) feature allows traffic to be classified into streams and given certain QoS treatment in accordance with defined per-hop behaviors. The switch supports both IPv4 and IPv6 packet classification.

1.3.5. Class of Service (CoS)

The Class of Service (CoS) queueing feature lets you directly configure certain aspects of switch queuing. This provides the desired QoS behavior for different types of network traffic when the complexities of DiffServ are not required. CoS queue characteristics, such as minimum guaranteed bandwidth and transmission rate shaping, are configurable at the queue (or port) level.

1.4. Management Features

1.4.1. Management Options

You can use the following methods to manage the switch:

- Use a telnet client, SSH client, or a direct console connection to access the CLI. The CLI syntax and semantics conform as much as possible to common industry practice.
- Use a network management system (NMS) to manage and monitor the system through SNMP. The switch supports SNMP v1/v2c/v3 over the UDP/IP transport protocol.

1.4.2. Management of Basic Network Information

The DHCP client on the switch allows the switch to acquire information such as the IP address and default gateway from a network DHCP server. You can also disable the DHCP client and configure static network information. Other configurable network information includes a Domain Name Server (DNS), host name to IP address mapping, and a default domain name.

The switch also includes a DHCPv6 client for acquiring IPv6 addresses, prefixes, and other IPv6 network configuration information.

1.4.3. File Management

You can upload and download files such as configuration files and system images by using TFTP, Secure FTP (SFTP), or Secure Copy (SCP). Configuration file uploads from the switch to a server are a good way to back up the switch configuration. You can also download a configuration file from a server to the switch to restore the switch to the configuration in the downloaded file.

1.4.4. Malicious Code Detection

This feature provides a mechanism to detect the integrity of the image, if the software binary is corrupted or tampered with while end user attempts to download the software image to the switch. This release addresses this problem by using digital signatures to verify the integrity of the binary image. It also provides flexibility to download a digitally signed configuration script and verify the digital signature to ensure the integrity of the downloaded configuration file.

1.4.5. Automatic Installation of Firmware and Configuration

The Auto Install feature allows the switch to upgrade the configuration file automatically during device initialization with limited administrative configuration on the device. The switch can obtain the necessary information from a DHCP server on the network.

1.4.6. Warm Reboot

The Warm Reboot feature reduces the time it takes to reboot the switch thereby reducing the traffic disruption in the network during a switch reboot. For a typical switch, the traffic disruption is reduced from about two minutes for a cold reboot to about 20 seconds for a warm reboot.

1.4.7. SNMP Alarms and Trap Logs

The system logs events with severity codes and timestamps. The events are sent as SNMP traps to a trap recipient list.

1.4.8. Remote Monitoring (RMON)

RMON is a standard Management Information Base (MIB) that defines current and historical MAC-layer statistics and control objects, allowing real-time information to be captured across the entire network. The data collected is defined in the RMON MIB, RFC 2819 (32-bit counters), RFC 3273 (64-bit counters), and RFC 3434 (High Capacity Alarm Table).

1.4.9. Statistics Application

The statistics application collects the statistics at a configurable time interval. The user can specify the port number(s) or a range of ports for statistics to be displayed. The configured time interval applies to all ports. Detailed statistics are collected between the specified time range in date and time format. The time range can be defined as having an absolute time entry and/or a periodic time. For example, a user can specify the statistics to be collected and displayed between 9:00 12 NOV 2011 (START) and 21:00 12 NOV 2011 (END) or schedule it on every MON, WED and FRI 9:00 (START) to 21:00 (END).

The user receives these statistics in a number of ways as listed below:

- User requests through CLI for a set of counters.
- User can configure the device to display statistics using syslog or email alert. The syslog or email alert messages are sent by statistics application at END time.

Note: The statistics are presented on the console at END time.

1.4.10. Log Messages

The switch maintains in-memory log messages as well as persistent logs. You can also configure remote logging so that the switch sends log messages to a remote log server. You can also configure the switch to send log messages to a configured SMTP server. This allows you to receive the log message in an e-mail account of your choice. Switch auditing messages, CLI command logging, and SNMP logging can be enabled or disabled.

1.4.11. System Time Management

The switch will obtain the system time and date through NTP (Network Time Protocol) service of Linux server, or you can set the time and date locally or configure the time zone on the switch via Linux.

1.4.12. Source IP Address Configuration

Syslog, TACACS, SNTP, sFlow, SNMP Trap, RADIUS, and DNS Clients allow the IP Stack to select the source IP address while generating the packet. This feature provides an option for the user to select an interface for the source IP address while the management protocol transmits packets to management stations. The source address is specified for each protocol.

1.4.13. Multiple Linux Routing Tables

On Linux systems, local and default IPv4 routes for the service port and network port are installed in routing tables dedicated to each management interface. Locally-originated IPv4 packets use these routing tables when the source IP address of the packet matches an address on one of these interfaces. This feature allows the Linux IP stack to use default routes for different interfaces simultaneously.

1.4.14. Open Network Install Environment Support

Open Network Install Environment (ONIE) allows customers to install their choice of network operating system (NOS) onto a switch. When the switch boots, ONIE enables the switch to fetch a NOS stored on a remote server. The remote server can hold multiple NOS images, and you can specify which NOS to load and run on the switch. ONIE support in the switch software facilitates automated data center provisioning by enabling a bare-metal network switch ecosystem.

ONIE is a small operating system. It is preinstalled as firmware and requires an ONIE-compliant boot loader (U-Boot/BusyBox), a kernel (Linux) and the ONIE discovery and execution application. For more information about ONIE, see http://onie.github.io/onie.

1.4.15. Interface Error Disable and Auto Recovery

If the switch detects an error condition for an interface, it places the interface in the diagnostic disabled state by shutting down the interface. The error-disabled interface does not allow any traffic until it is reenabled. You

can manually reenable the interface, or, if the Auto Recovery feature is enabled, the interface can be reenabled automatically after a configurable time-out period.

There are multiple reasons that may cause the switch to place an interface in the error-disabled state. Auto Recovery can be configured to take effect if an interface is error-disabled for any reason, or for some reasons but not others.

1.4.16. CLI Scheduler

The CLI scheduler allows customers to schedule fully-qualified EXEC mode CLI commands to run once, at specified intervals, at specified calendar dates and times, or upon system startup.

CLI scheduler has two basic processes. A policy list is configured containing lines of fully-qualified EXEC CLI commands to be run at the same time or same interval. One or more policy lists are then scheduled to run after a specified interval of time, at a specified calendar date and time, or upon system startup. Each scheduled occurrence can be set to run either once only or on a recurring basis.

1.5. Routing Features

1.5.1. IP Unnumbered

Each routing interface can be configured to borrow the IP address from the loopback interfaces and use this IP for all routing activities.

The IP Unnumbered feature was initially developed to avoid wasting an entire subnet on point-to-point serial links.

The IP Unnumbered feature can also be used in situations where adjacencies are transient and adjacent interfaces cannot be easily configured with IPv4 addresses in the same subnet. It also helps in reducing the configuration overhead in large scale Data-Center deployments.

1.5.2. Open Shortest Path First (OSPF)

Open Shortest Path First (OSPF) is a dynamic routing protocol commonly used within medium-to-large enterprise networks. OSPF is an interior gateway protocol (IGP) that operates within a single autonomous system.

1.5.3. Border Gateway Protocol (BGP)

BGP is an exterior routing protocol used in large-scale networks to transport routing information between autonomous systems (AS). As an interdomain routing protocol, BGP is used when AS path information is required to provide partial or full Internet routing downstream. The switch supports BGP version 4.

The following BGP features are supported:

- Proprietary BGP MIB support for reporting status variables and internal counters.
- Additional route map support:
    - Match as-path
    - Set as-path
    - Set local-preference
    - Set metric
- Supports for inbound and outbound neighbor-specific route maps.
- Handles the BGP RTO full condition.
- Supports for the show ip bgp command.
- Supports for the show ip bgp traffic command.
- Supports for the bgp always-compare-med command.
- Supports for the maximum number of BGP neighbors: 128.
- A prefix list is supported to filter the output of the show ip bgp command.
- Configurable maximum length of a received AS_PATH.
- Show command to list the routes accepted from a specific neighbor.
- Show command to list the routes rejected from a specific neighbor.
- Supports for BGP communities.
- Supports for IPv6.
- IPv6 Transport and Prefix list

Supports for BGP peer templates to simplify neighbor configuration.

1.5.4. VLAN Routing

The switch supports VLAN routing. You can also configure the software to allow traffic on a VLAN to be treated as if the VLAN were a router port.

1.5.5. IP Configuration

The switch IP configuration settings to allow you to configure network information for VLAN routing interfaces such as IP address and subnet mask, MTU size, and ICMP redirects. Global IP configuration settings for the switch allow you to enable or disable the generation of several types of ICMP messages and enable or disable the routing mode.

1.5.6. Address Resolution Protocol (ARP) Table Management

You can create static ARP entries and manage many settings for the dynamic ARP table, such as age time for entries, retries, and cache size.

1.5.7. BOOTP/DHCP Relay Agent

The switch BOOTP/DHCP Relay Agent feature relays BOOTP and DHCP messages between DHCP clients and DHCP servers that are located in different IP subnets.

1.5.8. IP Helper and UDP Relay

The IP Helper and UDP Relay features provide the ability to relay various protocols to servers on a different subnet.

1.5.9. Routing Table

The routing table displays information about the routes that have been dynamically learned. You can configure static and default routes and route preferences. A separate table shows the routes that have been manually configured.

1.5.10. Virtual Router Redundancy Protocol (VRRP)

VRRP provides hosts with redundant routers in the network topology without any need for the hosts to reconfigure or know that there are multiple routers. If the primary (master) router fails, a secondary router assumes control and continues to use the virtual router IP (VRIP) address.

VRRP Route Interface Tracking extends the capability of VRRP to allow tracking of specific route/interface IP states within the router that can alter the priority level of a virtual router for a VRRP group.

1.5.11. Algorithmic Longest Prefix Match (ALPM)

Algorithmic Longest Prefix Match (ALPM) is a protocol used by routers to select an entry from a forwarding table. When an exact match is not found in the forwarding table, the match with the longest subnet mask, also called longest prefix match, is chosen. It is called the longest prefix match because it is also the entry where the largest number of leading address bits of the destination address match those in the table entry.

ALPM enables support for large number of routes. (For BGP, 32k IPv4 routes and 24k IPv6 are supported.)

The SDM template, “dual-ipv4-and-ipv6 alpm” is available to accommodate a large number of routes.

1.5.12. Bidirectional Forwarding Detection

Bidirectional Forwarding Detection (BFD) is presented as a service to its user applications, providing the options to create and destroy a session with a peer device and reporting upon the session status. On the

switch, OSPF and BGP can use BFD for monitoring of their neighbors' availability in the network and for fast detection of connection faults with them.

1.5.13. VRF Lite Operation and Configuration

The Virtual Routing and Forwarding feature enables a router to function as multiple routers. Each virtual router manages its own routing domain, with its own IP routes, routing interfaces, and host entries. Each virtual router makes its own routing decisions, independent of other virtual routers. More than one virtual routing table may contain a route to a given destination. The network administrator can configure a subset of the router's interfaces to be associated with each virtual router. The router routes packets according to the virtual routing table associated with the packet's ingress interface. Each interface can be associated with at most one virtual router.

1.6. Layer 3 Multicast Features

1.6.1. Internet Group Management Protocol

The Internet Group Management Protocol (IGMP) is used by IPv4 systems (hosts and routers) to report their IP multicast group memberships to any neighboring multicast routers. The switch performs the “multicast router part” of the IGMP protocol, which means it collects the membership information needed by the active multicast router.

1.6.2. Protocol Independent Multicast

1.6.2.1. Spare Mode (PIM-SM)

Protocol Independent Multicast-Sparse Mode (PIM-SM) is used to efficiently route multicast traffic to multicast groups that may span wide area networks, and where bandwidth is a constraint. PIM-SM uses shared trees by default and implements source-based trees for efficiency. This data threshold rate is used to toggle between trees.

1.6.2.2. Source Specific Multicast (PIM-SSM)

Protocol Independent Multicast—Source Specific Multicast (PIM-SSM) is a subset of PIM-SM and is used for one-to-many multicast routing applications, such as audio or video broadcasts. PIM-SSM does not use shared trees.

1.6.2.3. PIM IPv6 Support

PIM-DM and PIM-SM support IPv6 routes.

1.6.3. MLD/MLDv2 (RFC2710/RFC3810)

MLD is used by IPv6 systems (listeners and routers) to report their IP multicast addresses memberships to any neighboring multicast routers. The implementation of MLD v2 is backward compatible with MLD v1.

MLD protocol enables the IPv6 router to discover the presence of multicast listeners, the nodes that want to receive the multicast data packets, on its directly attached interfaces. The protocol specifically discovers which multicast addresses are of interest to its neighboring nodes and provides this information to the multicast routing protocol that make the decision on the flow of the multicast data packets.

1.7. Data Center Features

1.7.1. Priority-Based Flow Control

The Priority-Based Flow Control (PFC) feature allows the user to pause or inhibit transmission of individual priorities within a single physical link. By configuring PFC to pause a congested priority (priorities) independently, protocols that are highly loss sensitive can share the same link with traffic that has different loss tolerances. Priorities are differentiated by the priority field of the 802.1Q VLAN header.

An interface that is configured for PFC is automatically disabled for 802.3x flow control.

1.7.2. Data Center Bridging Exchange Protocol

The Data Center Bridging Exchange Protocol (DCBX) is used by data center bridge devices to exchange configuration information with directly-connected peers. The protocol is also used to detect misconfiguration of the peer DCBX devices and optionally, for configuration of peer DCBX devices.

1.7.3. CoS Queuing and Enhanced Transmission Selection

The CoS Queuing feature allows the switch administrator to directly configure certain aspects of the device hardware queuing to provide the desired QoS behavior for different types of network traffic. The priority of a packet arriving at an interface can be used to steer the packet to the appropriate outbound CoS queue through a mapping table. CoS queue characteristics such as minimum guaranteed bandwidth, transmission rate shaping, etc. are user configurable at the queue (or port) level.

Enhanced Transmission Selection (ETS) allows Class of Service (CoS) configuration settings to be advertised to other devices in a data center network through DCBX ETS TLVs. CoS information is exchanged with peer DCBX devices using ETS TLVs.

1.7.4. VXLAN Gateway

Logically segregated virtual networks in a data center are sometimes referred to as data center VPNs. The VXLAN Gateway is a solution that allows VXLAN to communicate with another network, particularly a VLAN. It offers VXLAN Tunnel Endpoint (VTEP) functionality for VXLAN tunnels on the switch.

VXLAN is a layer-3 function, IP-based technologies that prepend an existing layer-2 frame with a new IP header, providing layer-3 based tunneling capabilities for layer-2 frames. This essentially enables a layer-2 domain to extend across a layer-3 boundary.

For the traffic from a VXLAN to use services on physical devices in a distant network, the traffic must pass through a VXLAN Gateway.

The VXLAN Gateway feature is configurable through the CLI. It also offers an Overlay API to facilitate programming from external agents.

2. Getting Started

2.1. Accessing the switch Command-Line Interface

The command-line interface (CLI) provides a text-based way to manage and monitor the switch features. You can access the CLI by using a direct connection to the console port or by using a Telnet or SSH client.

To access the switch by using Telnet or Secure Shell (SSH), the switch must have an IP address configured on either the service port or the management VLAN interface, and the management station you use to access the device must be able to ping the switch IP address. DHCP is enabled by default on the service port. It is disabled on the management VLAN interface.

Note: For information about changing the default settings for Telnet and SSH access methods, see “Configuring and Applying Authentication Profiles”.

2.1.1. Connecting to the Switch Console

To connect to the switch and configure or view network information, use the following steps:

1. Using a straight-through modem cable, connect a VT100/ANSI terminal or a workstation to the console (serial) port.

If you attached a PC, Apple, or UNIX workstation, start a terminal-emulation program, such as

HyperTerminal or TeraTerm.

2. Configure the terminal-emulation program to use the following settings:

- Baud rate: 115200 bps
- Data bits: 8
- Parity: none
- Stop bit: 1
- Flow control: none

3. Power on the switch:

After the system completes the boot cycle, the switch login prompt appears.

2.1.2. Login User ID and Password

You can log in to the switch using any of the following methods:

- Serial Console
- SSH (Secure Shell)
- Telnet, using special port 1224

The default login name is **admin** and the default password is **EndGame**.

After you are log in to the switch, to access the switch CLI, you must provide a user name and password. The default user name is **admin**, but the first time that you access the switch CLI, no default password is required. That is, just press the **Enter** button. When you log in for the first time, the switch CLI prompts you to change the switch CLI password.

The switch CLI lets you access all switch configuration commands.

When you access the switch using the serial console or SSH method, press **<Ctrl> + z** or enter **logout** at the switch CLI prompt to exit the switch CLI and display the following menu options.

```
====================================
NETGEAR M4500 Menu
====================================
1: CLI Console
2: Firmware update using SCP
3: Firmware update using TFTP
4: Reboot
====================================

Enter your menu option:
```

Enter menu option 1 to access the switch CLI.

2.1.3. Accessing the Switch CLI through the Network

Note: The telnet port number is **1223**. The SSH port number is **1234**.

Remote management of the switch is available through the service port or through the management VLAN interface. To use telnet, SSH, or SNMP for switch management, the switch must be connected to the network, and you must know the IP or IPv6 address of the management interface. The switch has no IP address by default. The DHCP client on the service port is enabled, and the DHCP client on the management VLAN interface is disabled.

After you configure or view network information, configure the authentication profile for telnet or SSH (see “Configuring and Applying Authentication Profiles”) and physically and logically connect the switch to the network, you can manage and monitor the switch remotely. You can also continue to manage the switch through the terminal interface via the console port.

2.1.4. Using the Service Port or Management VLAN Interface for Remote Management

The service port is a dedicated Ethernet port for out-of-band management. We recommend that you use the service port to manage the switch. Traffic on this port is segregated from operational network traffic on the switch ports and cannot be switched or routed to the operational network. Additionally, if the production network is experiencing problems, the service port still allows you to access the switch management interface and troubleshoot issues. Configuration options on the service port are limited, which makes it difficult to accidentally cut off management access to the switch.

Alternatively, you can choose to manage the switch through the production network, which is known as in-band management. Because in-band management traffic is mixed in with production network traffic, it is subject to all of the filtering rules usually applied on a switched/routed port such as ACLs and VLAN tagging. You can access the in-band network management interface through a connection to any front-panel port.

2.1.4.1. Configuring Service Port Information

To disable DHCP and manually assign an IPv4 address, enter:

```
serviceport protocol none
serviceport ip ipaddress netmask [gateway]
```

For example, `serviceport ip 192.168.2.23 255.255.255.0 192.168.2.1`

To disable DHCP and manually assign an IPv6 address and (optionally) default gateway, enter:

```
serviceport protocol none
serviceport ipv6 address address/prefix-length [eui64]
serviceport ipv6 gateway gateway
```

To view the assigned or configured network address, enter:

```
show serviceport
```

To enable the DHCP client on the service port, enter:

```
serviceport protocol dhcp
```

2.1.4.2. Configuring the In-Band Management Interface

To use a DHCP server to obtain the IP address, subnet mask, and default gateway information, enter:

```
(Switch) (Config)#interface vlan 1
(Switch) (if-vlan1)#ip address dhcp
```

To manually configure the IPv4 address, subnet mask, and (optionally) default gateway, enter:

```
(Switch)(config)#interface vlan 1
(Switch)(if-vlan 1)#ip address ipaddress netmask
```

```
(Switch)(if-vlan 1)#exit
(Switch)(Config)#ip default-gateway gateway
```

For example:

```
(Switch)(if-vlan 1)#ip address 192.168.1.253 255.255.255.0
(Switch)(Config)#ip default-gateway 192.168.1.254
```

To manually configure the IPv6 address, subnet mask, enter:

```
(Switch)(config)#interface vlan 1
(Switch)(if-vlan 1)#ipv6 address address/prefix-length [eui64]
```

To view the In-Band management information, enter:

```
show ip interface.
show ipv6 interface
```

To save these changes so they are retained during a switch reset, enter the following command:

```
copy running-config startup-config
```

2.1.5. DHCP Option 61

DHCP Option 61 (client Identifier) allows the DHCP server to be configured to provide an IP address to a switch based on its Media Access Control (MAC) Address or an ID entered into the system. DHCP servers use this value to index their database of address bindings. This value is expected to be unique for all clients in an administrative domain. This option allows the system to move from one part of the network to another while maintaining the same IP address.

DHCP client Identifier (Option 61) is used by DHCP clients to specify their unique identifier. The client identifier option is optional and can be specified while configuring the DHCP on the interfaces. DHCP Option 61 is enabled by default.

2.1.5.1. Configuring DHCP Option 61

Configuring the DHCP with client-id (option 61) differs depending on the port or interface. See the following information:

Service Port:

To enable DHCP with client-id (option 61) on from the service port, issue the following command:

```
(Switch) #serviceport protocol dhcp client-id
```

In-Band management Port:

To enable DHCP with client-id (option 61) on from the In-Band management port, issue the following command:

```
(Switch) (Config)#interface vlan 1
(Switch) (if-vlan1)#ip address dhcp client-id
```

Routing Enabled Interface:

To enable DHCP with client-id (option 61) on from on the routing enabled interface, issue the following command

in interface configuration mode.

```
(Switch) (Interface 0/1)#ip address dhcp client-id
```

Physical Interface:

To enable DHCP with client-id (option 61) on from on the physical interface, issue the commands as shown below:

```
(Switch) #config
(Switch) (Config)#interface 0/4
(Switch) (Interface 0/4)#ip address dhcp client-id
```

VLAN Interface:

To enable DHCP with client-id (option 61) on from on the VLAN interface, issue the commands as shown below:

```
(Switch) #config
(Switch) (Config)#interface vlan 10
(Switch) (Interface vlan 10)#ip address dhcp client-id
```

2.2. Understanding the User Interfaces

The switch includes a set of comprehensive management functions for configuring and monitoring the system by using one of the following two methods:

- Command-Line Interface (CLI)
- Simple Network Management Protocol (SNMP)

These standards-based management methods allow you to configure and monitor the components of the software. The method you use to manage the system depends on your network size and requirements, and on your preference.

2.2.1. Using the Command-Line Interface

The command-line interface (CLI) is a text-based way to manage and monitor the system. You can access the CLI by using a direct serial connection or by using a remote logical connection with telnet or SSH.

The CLI groups commands into modes according to the command function. Each of the command modes supports specific software commands. The commands in one mode are not available until you switch to that particular mode, with the exception of the User EXEC mode commands. You can execute the User EXEC mode commands in the Privileged EXEC mode.

To display the commands available in the current mode, enter a question mark (?) at the command prompt. To display the available command keywords or parameters, enter a question mark (?) after each word you type at the command prompt. If there are no additional command keywords or parameters, or if additional parameters are optional, the following message appears in the output:

```
<cr>Press Enter to execute the command
```

For more information about the CLI, see the *M4500 Intelligent Fully Managed Switches CLI Command Reference Manual.*

The *M4500 Intelligent Fully Managed Switches CLI Command Reference Manual* lists each command available from the CLI by the command name and provides a brief description of the command. Each command reference also contains the following information:

- The command keywords and the required and optional parameters.
- The command mode you must be in to access the command.
- The default value, if any, of a configurable setting on the device.

The `show` commands in the document also include a description of the information that the command shows.

2.2.2. Using SNMP

SNMP is enabled by default. The `show sysinfo` command displays the information you need to configure an SNMP manager to access the switch. You can configure SNMP groups and users that can manage traps that the SNMP agent generates.

The switch uses both standard public MIBs for standard functionality and private MIBs that support additional switch functionality. All private MIBs begin with a "-" prefix. The main object for interface configuration is in -SWITCHING-MIB, which is a private MIB. Some interface configurations also involve objects in the public MIB, IF-MIB.

2.2.2.1. SNMPv3

SNMP version 3 (SNMPv3) adds security and remote configuration enhancements to SNMP. You can configure SNMP server, users, and traps for SNMPv3. Any user can connect to the switch using the SNMPv3 protocol, but for authentication and encryption, you need to configure a new user profile. To configure a profile by using the CLI, see the SNMP section in the *M4500 Intelligent Fully Managed Switches CLI Command Reference Manual*.

2.2.2.2. SNMP Configuration example

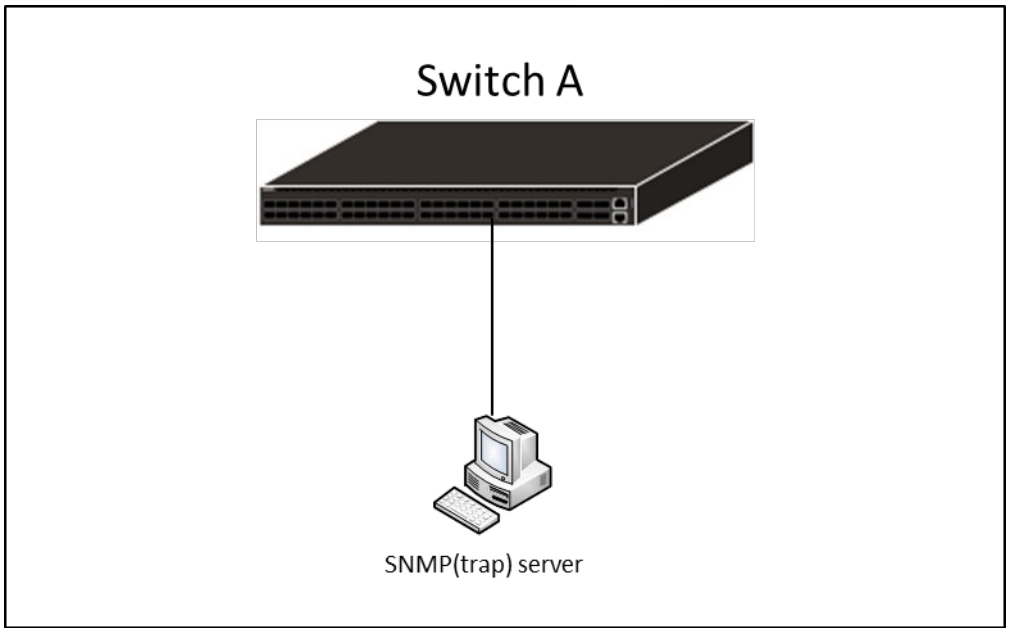


Figure 2-1: SNMP Configuration Topology

SNMPv1, v2c community and trap configuration example

1. Add new community testRO for read only and testRW for read-write.

```
(Switch) (Config)#snmp-server community testRO ro
(Switch) (Config)#snmp-server community testRW rw
```

2. Setup SNMP trap host IP address.

```
(Switch) (Config)#snmp-server host 172.16.1.100 traps version 1 testRO
(Switch) (Config)#snmp-server host 172.16.2.100 traps version 2 testRO
```

3. Verify the configuration.

```
(Switch) #show snmp
Community-String    Community-Access View Name       IP Address
------------------ ---------------- -------------- ----------------
private              Read/Write       Default          All
public               Read Only        Default          All
testRO               Read Only        Default          All
testRW               Read/Write       Default          All

Community-String    Group Name                    IP Address
------------------ ---------------------------- ----------------
private              DefaultWrite                  All
```

```
public               DefaultRead                   All
testRO               DefaultRead                   All
testRW               DefaultWrite                  All

Traps are enabled.
Authentication trap is enabled.

Version 1,2 notifications
Target Address   Type    Community            Version UDP    Filter   TO   Retries
                                                        Port    name    Sec
--------------- ------- ------------------- ------- ------ -------- --- ------
172.16.1.100     Trap    testRO                1       162
172.16.2.100     Trap    testRO                2       162

Version 3 notifications
Target Address   Type    Username           Security    UDP    Filter   TO   Retries
                                               Level    Port   name     Sec
---------------- ------- ----------------- -------- ------ -------- --- -------

System Contact:
System Location:
```

SNMPv3 configuration example

1. Configure a view and oid-tree included iso.

```
(Switch) (Config)#snmp-server view testVIEW iso included
```

2. Configure a group which use testVIEW for read, write and notify (trap).

```
(Switch) (Config)#snmp-server group testGROUP v3 noauth read testVIEW write testVIEW notify
testVIEW
```

3. Add a user named testUSER and assign to testGROUP.

```
(Switch) (Config)#snmp-server user testUSER testGROUP
```

4. Setup SNMPv3 trap host IP address.

```
(Switch) (Config)#snmp-server host 172.16.1.102 traps version 3 testUSER noauth
```

5. Verify the configuration.

```
(Switch) #show snmp views

          Name                          OID Tree                Type
-------------------------- ------------------------------ ----------
Default                       iso                            Included
Default                       snmpVacmMIB                    Excluded
Default                       usmUser                        Excluded
Default                       snmpCommunityTable             Excluded
testVIEW                      iso                            Included
DefaultSuper                  iso                            Included

(Switch) #show snmp group

       Name          Context              Security             Views
                      Prefix     Model     Level       Read    Write    Notify
------------------ ---------- ----- ------------- -------- -------- --------
testGROUP             ""         V3    NoAuth-NoPriv testVIEW testVIEW testVIEW
DefaultRead           ""         V1    NoAuth-NoPriv Default  ""       Default
DefaultRead           ""         V2    NoAuth-NoPriv Default  ""       Default
DefaultRead           ""         V3    NoAuth-NoPriv Default  ""       Default
DefaultRead           ""         V3    Auth-NoPriv   Default  ""       Default
DefaultRead           ""         V3    Auth-Priv     Default  ""       Default
DefaultSuper          ""         V1    NoAuth-NoPriv DefaultS DefaultS DefaultS
                                                     uper     uper     uper
DefaultSuper          ""         V2    NoAuth-NoPriv DefaultS DefaultS DefaultS
                                                     uper      uper     uper
DefaultSuper          ""         V3    NoAuth-NoPriv DefaultS DefaultS DefaultS
                                                     uper      uper     uper
DefaultWrite          ""         V1    NoAuth-NoPriv Default  Default  Default
DefaultWrite          ""         V2    NoAuth-NoPriv Default  Default  Default
DefaultWrite          ""         V3    NoAuth-NoPriv Default  Default  Default
DefaultWrite          ""         V3    Auth-NoPriv   Default  Default  Default
DefaultWrite          ""         V3    Auth-Priv     Default  Default  Default
```

```
(Switch) #show snmp user

     Name            Group Name     Auth Priv
                                     Meth Meth    Remote Engine ID
---------------- ------------- ---- ---- -------------------------
testUSER            testGROUP                    80001c4c03000000000004

(Switch) #show snmp

Community-String     Community-Access View Name        IP Address
------------------ --------------- ---------------- ----------------
private              Read/Write       Default          All
public               Read Only        Default          All
testRO               Read Only        Default          All
testRW               Read/Write       Default          All


Community-String     Group Name                     IP Address
------------------ ------------------------------ ----------------
private              DefaultWrite                   All
public               DefaultRead                    All
testRO               DefaultRead                    All
testRW               DefaultWrite                   All

Traps are enabled.
Authentication trap is enabled.

Version 1,2 notifications
Target Address   Type    Community           Version UDP   Filter   TO   Retries
                                                      Port   name    Sec
--------------- ------- ----------------- ------ ------ -------- --- ------
172.16.1.100     Trap    testRO               1       162
172.16.2.100     Trap    testRO               2       162
```

```
Version 3 notifications
Target Address   Type    Username        Security UDP    Filter   TO   Retries
                                            Level    Port   name     Sec
--------------- ------- -------------- -------- ------ -------- --- -------
172.16.1.102     Trap    testUSER        NoAuth-N  162             15  3

System Contact:
System Location:
```

3. Configuring L2 Switching Features

3.1. Port Configuration

You can configure the ports on the switch to support speeds from 1G/10G/40G to 25/50/100G. You can use the software to change a port from one mode to another. When the port is configured in 40/100Gbps mode, the four 10/25Gbps ports at the same physical interface are disabled.

3.1.1. 100G Port-mode Command

Use the port-mode command to configure a 25G/100G QSFP port in either 4x10/25G mode, 2x50G or 1x40/100G mode. This command can only be executed on the original 25G/100G interface. Entering the command on any other type of interface will give an error. This command does not operate in interface range mode.

3.1.1.1. 100G Port Mode Configuration Example

The following example guide you how to configure port 1 as 40G, port 2 fan-out to 4x25G and port 3 fan-out to 2x50G on 100G ports.

1. Enter interface mode of port 1, and configure to 40G port.

```
(Switch) (Config)#interface 0/1
(Switch) (Interface 0/1)#port-mode 1x40G
(Switch) (Interface 0/1)#exit
```

2. Enter interface mode of port 2, and configure to 4x25G port.

```
(Switch) (Config)#interface 0/2
(Switch) (Interface 0/2)#port-mode 4x25G
(Switch) (Interface 0/2)#exit
```

3. Enter interface mode of port 3, and configure to 2x50G port.

```
(Switch) (Config)#interface 0/3
(Switch) (Interface 0/3)#port-mode 2x50G
(Switch) (Interface 0/3)#exit
```

4. Using show interface port-mode to check the hardware profile information for the all ports.

```
(Switch) #show interface port-mode

100G/40G       Configured  Operating   Expandable  Expanded
Interface      Mode        Mode        Option(s)   Interfaces
-------------------------------------------------------------
```

```
0/1             1x40G       1x40G       4x10G       0/33-36
                                        4x25G       0/33-36
                                        2x50G       0/161-162
0/2             4x25G       4x25G       4x10G       0/37-40
                                        4x25G       0/37-40
                                        2x50G       0/163-164
0/3             2x50G       2x50G       4x10G       0/41-44
                                        4x25G       0/41-44
                                        2x50G       0/165-166
0/4             1x100G      1x100G      4x10G       0/45-48
                                        4x25G       0/45-48
                                        2x50G       0/167-168
...
```

3.2. Virtual Local Area Networks

By default, all switchports on the switch are in the same broadcast domain. This means when one host connected to the switch broadcasts traffic, every device connected to the switch receives that broadcast. All ports in a broadcast domain also forward multicast and unknown unicast traffic to the connected host. Large broadcast domains can result in network congestion, and end users might complain that the network is slow. In addition to latency, large broadcast domains are a greater security risk since all hosts receive all broadcasts.

Virtual Local Area Networks (VLANs) allow you to divide a broadcast domain into smaller, logical networks. Like a bridge, a VLAN switch forwards traffic based on the Layer 2 header, which is fast, and like a router, it partitions the network into logical segments, which provides better administration, security, and management of multicast traffic.

Network administrators have many reasons for creating logical divisions, such as department or project membership. Because VLANs enable logical groupings, members do not need to be physically connected to the same switch or network segment. Some network administrators use VLANs to segregate traffic by type so that the time-sensitive traffic, like voice traffic, has priority over other traffic, such as data. Administrators also use VLANs to protect network resources. Traffic sent by authenticated clients might be assigned to one VLAN, while traffic sent from unauthenticated clients might be assigned to a different VLAN that allows limited network access.

When one host in a VLAN sends a broadcast, the switch forwards traffic only to other members of that VLAN. For traffic to go from a host in one VLAN to a host in a different VLAN, the traffic must be forwarded by a layer 3 device, such as a router. VLANs work across multiple switches, so there is no requirement for the hosts to be located near each other to participate in the same VLAN.

Note: The switch supports VLAN routing. When you configure VLAN routing, the switch acts as a layer 3 device and can forward traffic between VLANs. For more information, see “VLAN Routing”.

Each VLAN has a unique number, called the VLAN ID. The switch supports a configurable VLAN ID range of 2–4093. A VLAN with VLAN ID 1 is configured on the switch by default. You can associate a name with the VLAN

ID. In a tagged frame, the VLAN is identified by the VLAN ID in the tag. In an untagged frame, the VLAN identifier is the Port VLAN ID (PVID) specified for the port that received the frame. For information about tagged and untagged frames, see “VLAN Tagging”.

You can add individual ports and Port-channels as VLAN members.

The following figure shows an example of a network with three VLANs that are department-based. The file server and end stations for the department are all members of the same VLAN.

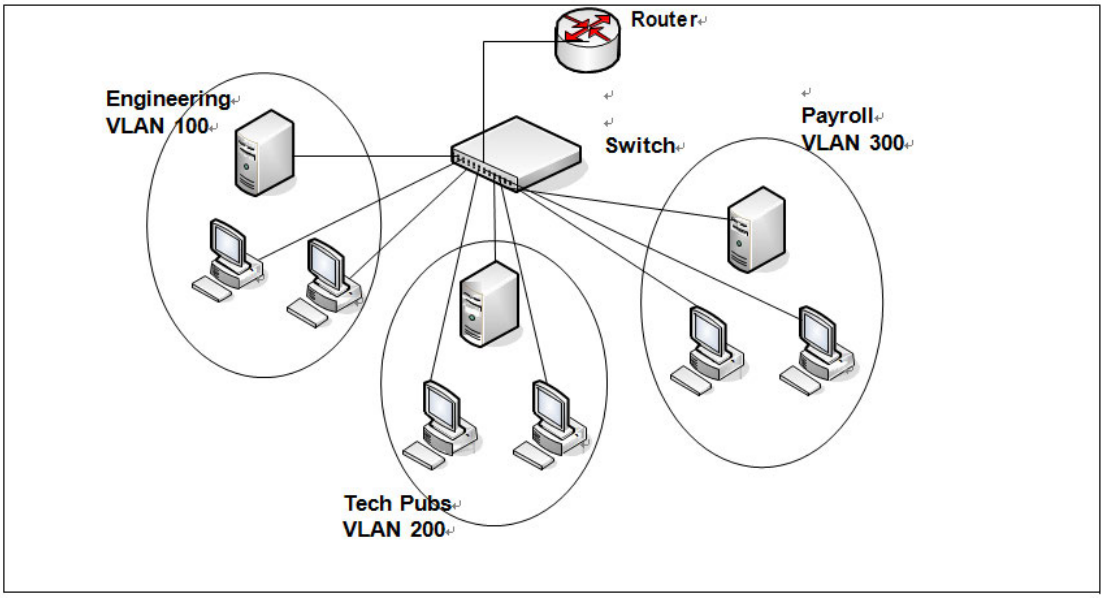


Figure 3-1: Simple VLAN Topology

In this example, each port is manually configured so that the end station attached to the port is a member of the VLAN configured for the port. The VLAN membership for this network is port-based or static.

3.2.1. VLAN Tagging

The switch supports IEEE 802.1Q tagging. Ethernet frames on a tagged VLAN have a 4-byte VLAN tag in the header. VLAN tagging is required when a VLAN spans multiple switches, which is why trunk ports transmit and receive only tagged frames.

Tagging may be required when a single port supports multiple devices that are members of different VLANs. For example, a single port might be connected to an IP phone, a PC, and a printer (the PC and printer are connected via ports on the IP phone). IP phones are typically configured to use a tagged VLAN for voice traffic, while the PC and printers typically use the untagged VLAN.

When a port is added to a VLAN as an untagged member, untagged packets entering the switch are tagged with the PVID (also called the *native VLAN*) of the port. If the port is added to a VLAN as an untagged member, the port does not add a tag to a packet in that VLAN when it exits the port. Configuring the PVID for an interface is useful when untagged and tagged packets will be sent and received on that port and a device connected to the interface does not support VLAN tagging.

When ingress filtering is on, the frame is dropped if the port is not a member of the VLAN identified by the VLAN ID in the tag. If ingress filtering is off, all tagged frames are forwarded. The port decides whether to forward or drop the frame when the port receives the frame.

3.2.2. Double-VLAN Tagging

For trunk ports, which are ports that connect one switch to another switch, the switch supports double-VLAN tagging. This feature allows service providers to create Virtual Metropolitan Area Networks (VMANs). With double-VLAN tagging, service providers can pass VLAN traffic from one customer domain to another through a metro core in a simple and cost-effective manner. By using an additional tag on the traffic, the switch can differentiate between customers in the MAN while preserving an individual customer's VLAN identification when the traffic enters the customer's 802.1Q domain.

With the introduction of this second tag, customers are no longer required to divide the 4-byte VLAN ID space to send traffic on an Ethernet-based MAN. That is, every frame that is transmitted from an interface has a double- VLAN tag attached, while every packet that is received from an interface has a tag removed (if one or more tags are present).

In the following figure, two customers share the same metro core. The service provider assigns each customer a unique ID so that the provider can distinguish between the two customers and apply different rules to each. When the configurable EtherType is assigned to something different than the 802.1Q (0x8100) EtherType, it allows the traffic to have added security from misconfiguration while exiting the metro core. For example, if the edge device on the other side of the metro core is not stripping the second tag, the packet would never be classified as an 802.1Q tag, so the packet would be dropped rather than forwarded in the incorrect VLAN.

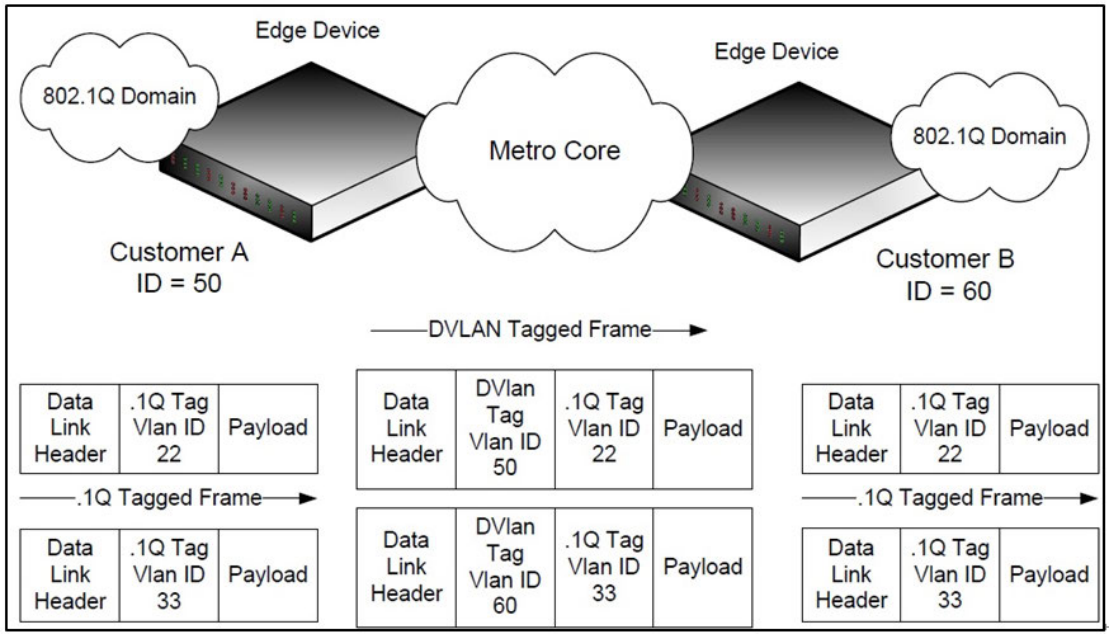


Figure 3-2: Double VLAN Tagging Network Example

3.2.3. Default VLAN Behavior

One VLAN exists on the switch by default. The VLAN ID is 1, and all ports are included in the VLAN as access ports, which are untagged. This means when a device connects to any port on the switch, the port forwards the packets without inserting a VLAN tag. If a device sends a tagged frame to a port, the frame is dropped. Since all ports are members of this VLAN, all ports are in the same broadcast domain and receive all broadcast and multicast traffic received on any port.

When you add a new VLAN to the VLAN database, no ports are members.
The configurable VLAN range is 2–4093. VLANs 4094 and 4095 are reserved.

The following table shows the default values or maximum values for VLAN features.

Feature	*Value*
Default VLAN ID	1
VLAN Name	default
VLAN Range	2–4093
Frames accepted	Untagged Incoming untagged fxrames are classified into the VLAN whose VLAN ID is the currently configured PVID.
Frames sent	Untagged
Ingress Filtering	On
PVID	1
Double-VLAN tagging	Disabled If double-VLAN tagging is enabled, the default EtherType value is 802.1Q

Table 3-1: VLAN default and maximum values

3.2.4. VLAN Configuration Example

A network administrator wants to create the VLANs that are listed in the following table.

VLAN ID	*VLAN Name*	*VLAN Type*	*Purpose*
100	Engineering	Port-based	All employees in the Engineering department use this VLAN. Confining this department's traffic to a single VLAN helps reduce the amount of traffic in the broadcast domain, which increases bandwidth.
200	Marketing	Port-based	All employees in the Marketing department use this VLAN.
300	Payroll	Port-based	The payroll department has sensitive traffic and needs its own VLAN to help keep that traffic private.

Table 3-2: Example VLAN

The following figure shows the network topology for this example. As the figure shows, there are two switches, two file servers, and many hosts. One switch has an uplink port that connects it to a layer 3 device and the rest of the corporate network.

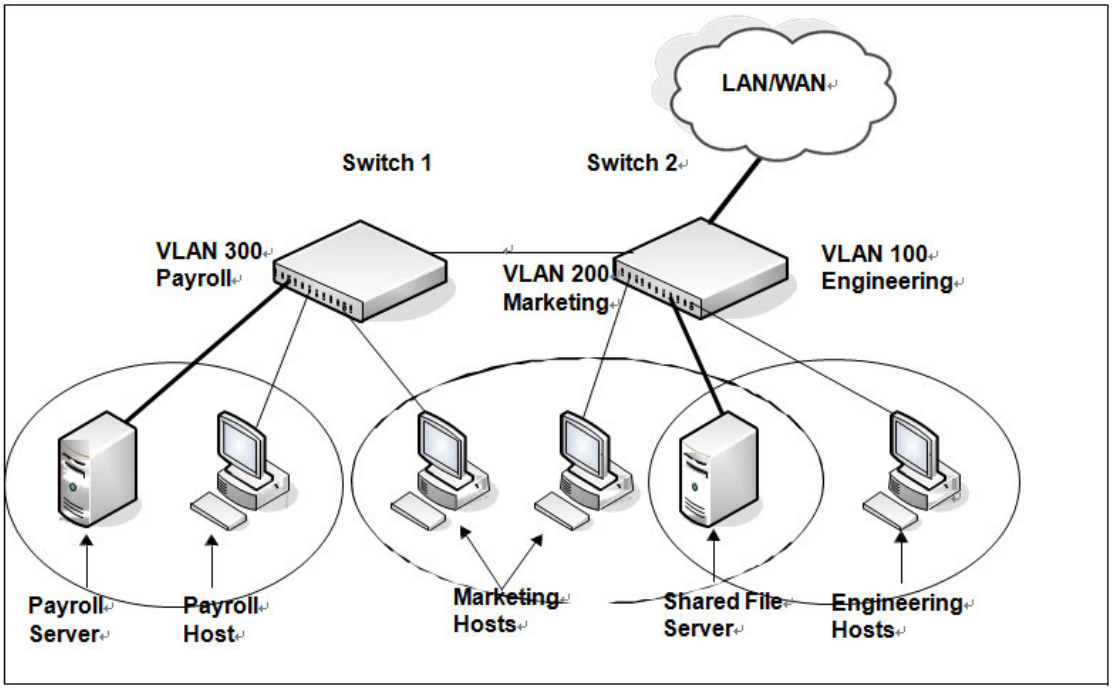


Figure 3-3: Network Topology for VLAN Configuration

The network in the previous figure has the following characteristics:

•Each connection to a host represents multiple ports and hosts.

•The Payroll and File servers are connected to the switches through a Port-channel.

•Some of the Marketing hosts connect to Switch 1, and some connect to Switch 2.

•The Engineering and Marketing departments share the same file server.

•Because security is a concern for the Payroll VLAN, the ports and Port-channel that are members of this VLAN will accept and transmit only traffic tagged with VLAN 300.

The following table shows the port assignments on the switches.

Port/Port-channel	*Function*
Switch 1	
1	Connects to Switch 2
2–15	Host ports for Payroll
16–20	Host ports for Marketing
Port-channel 1 (ports 21–24)	Connects to Payroll server
Switch 2	
1	Connects to Switch 1
2–10	Host ports for Marketing
11–30	Host ports for Engineering
Port-channel 1 (ports 35–39)	Connects to file server
Port-channel 2 (ports 40–44)	Uplink to router

3.2.4.1. Configuring the VLANs and Ports on Switch 1

Use the following steps to configure the VLANs and ports on Switch 1. None of the hosts that connect to Switch 1 use the Engineering VLAN (VLAN 100), so it is not necessary to create it on that switch. To configure Switch 1:

1. Create VLANs 200 (Marketing), 300 (Payroll), and associate the VLAN ID with the appropriate name.

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 200,300
(Switch) (Vlan)#vlan name 200 Marketing (Switch) (Vlan)#vlan name 300 Payroll
(Switch) (Vlan)#exit
```

2. Assign ports 16–20 to the Marketing VLAN.

```
(Switch) #configure
(Switch) (Config)#interface range 0/16-0/20
(Switch) (Interface 0/16-0/20)#switchport allowed vlan add 200
(Switch) (Interface 0/16-0/20)#switchport native vlan 200
(Switch) (Interface 0/16-0/20)#exit
```

3. Assign ports 2–15 to the Payroll VLAN

```
(Switch) (Config)#interface range 0/2-0/15
(Switch) (Interface 0/2-0/15)#switchport allowed vlan add 300
(Switch) (Interface 0/2-0/15)#switchport native vlan 300
(Switch) (Interface 0/2-0/15)#exit
```

4. Assign Port-channel1 to the Payroll VLAN and configure the frames to always be transmitted tagged with a PVID of 300.

```
(Switch) (Config)#interface port-channel 1
(Switch) (if-port-channel ch1)#switchport allowed vlan add tagged 300
(Switch) (if-port-channel ch1)#switchport native vlan 300
(Switch) (if-port-channel ch1)#exit
```

5. Configure port 1 as a trunk port and add VLAN 200 and VLAN 300 as members. Trunk ports accept and transmits tagged frames only and have ingress filtering enabled.

```
(Switch) (Config)#interface 0/1
(Switch) (Interface 0/1)#switchport acceptable-frame-types tagged
(Switch) (Interface 0/1)#switchport allowed vlan add tagged 200,300
(Switch) (Interface 0/1)#switchport ingress-filtering
(Switch) (Interface 0/1)#exit
```

```
(Switch) (Config)#exit
```

6. To save the configuration so that it persists across a system reset, use the following command:

```
(Switch) #copy running-config startup-config
```

7. View the VLAN settings.

```
(Switch) #show vlan
VLAN ID VLAN Name                         VLAN Type         Interface(s)
------- --------------------------------- ---------- -------------------------
1       default                           Default    0/1,0/2,0/3,0/4,0/5,0/6,
                                                     0/7,0/8,0/9,0/10,0/11,
                                                     0/12,0/13,0/14,0/15,0/16,
                                                     ...
200     Marketing                         Static     0/1,0/16,0/17,0/18,0/19,
                                                     0/20
300     Payroll                           Static     0/1,0/2,0/3,0/4,0/5,0/6,
                                                     0/7,0/8,0/9,0/10,0/11,
                                                     0/12,0/13,0/14,0/15,ch1
(Switch) #show vlan id 300
VLAN ID: 300
VLAN Name: Payroll
VLAN Type: Static



 Interface   Current   Configured   Tagging
 ----------  --------  -----------  --------
 0/1         Include   Include      Tagged
 0/2         Include   Include      Untagged
 0/3         Include   Include      Untagged
 0/4         Include   Include      Untagged
 0/5         Include   Include      Untagged
--More-- or (q)uit
```

8. View the VLAN information for a port.

```
(Switch) #show interface switchport 0/1
Interface..................................... 0/1
Native VLAN................................... 1
Mode.......................................... General
Ingress Filtering............................. Enable
```

```
Acceptable Frame Type..................... VLAN Only

Interface is member in:

VLAN ID VLAN Name                         VLAN Type         Egress rule
------- --------------------------------- ----------------- -----------
1       default                           Default           Untagged
200     Marketing                         Static            Tagged
300     Payroll                           Static            Tagged
```

3.2.4.2. Configuring the VLANs and Ports on Switch 2

Use the following steps to configure the VLANs and ports on Switch 2. Many of the procedures in this section are the same as procedures used to configure Switch 1. For more information about specific procedures, see the details and figures in the previous section.

To configure Switch 2:

1. Create the Engineering, Marketing, and Payroll VLANs.

Although the Payroll hosts do not connect to this switch, traffic from the Payroll department must use Switch 2 to reach the rest of the network and Internet through the uplink port. For that reason, Switch 2 must be aware of VLAN 300 so that traffic is not rejected by the trunk port.

2. Configure ports 2-10 to participate in VLAN 200.

3. Configure ports 11–30 to participate in VLAN 100.

4. Configure Port-channel 1 to participate in VLAN 100 and VLAN 200.

5. Configure port 1 and Port-channel 2 as participants in ports and add VLAN 100, VLAN 200, and VLAN 300 that accept and transit tagged frames only.

6. Enable ingress filtering on port 1 and Port-channel 2.

7. If desired, copy the running configuration to the startup configuration.

8. View VLAN information for the switch and ports.

3.3. Switchport Modes

You can configure each port on the switch to be in one of the following modes:

•**Access**—Access ports are intended to connect end-stations to the system, especially when the end- stations are incapable of generating VLAN tags. Access ports support a single VLAN (the PVID). Packets received

untagged are processed as if they are tagged with the access port PVID. Packets received that are tagged with the PVID are also processed. Packets received that are tagged with a VLAN other than the PVID are dropped. If the VLAN associated with an access port is deleted, the PVID of the access port is set to VLAN 1. VLAN 1 may not be deleted.

•**Trunk**—Trunk-mode ports are intended for switch-to-switch links. Trunk ports can receive both tagged and untagged packets. Tagged packets received on a trunk port are forwarded on the VLAN contained in the tag if the trunk port is a member of the VLAN. Untagged packets received on a trunk port are forwarded on the native VLAN. Packets received on another interface belonging to the native VLAN are transmitted untagged on a trunk port.

•**General**—General ports can act like access or trunk ports or a hybrid of both. VLAN membership rules that apply to a port are based on the switchport mode configured for the port.

The following table shows the behavior of the three switchport modes.

Mode	*VLAN Membership*	*Frames Accepted*	*Frames Sent*	*Ingress Filtering*
Access	One VLAN	Untagged/Tagged	Untagged	Always On
Trunk	All VLANs that exist in the system (default)	Untagged/Tagged	Tagged and Untagged	Always On
General	As many as desired	Tagged or Untagged	Tagged or Untagged	On or Off

Table 3-4: Switchport Mode Behavior

When a port is in General mode (by default all interfaces are in general mode), all VLAN features are configurable. When ingress filtering is on, the frame is dropped if the port is not a member of the VLAN identified by the VLAN ID in the tag. If ingress filtering is off, all tagged frames are forwarded. The port decides whether to forward or drop the frame when the port receives the frame.

The following example configures a port in Access mode with a single VLAN membership in VLAN 10:

```
(Switch) #config
(Switch) (Config)#interface 0/5
(Switch) (Interface 0/5)#switchport mode access
(Switch) (Interface 0/5)#switchport access vlan 10
(Switch) (Interface 0/5)#exit
```

The following example configures a port in Trunk mode. The **switchport trunk allowed vlan** command with the **add** keyword adds the list of VLANs that can receive and send traffic on the interface in tagged format when in trunking mode. Alternatively, the all keyword can be used to specify membership in all VLANs, the **remove** keyword can be used to remove membership. If this command is omitted, the port is a member of all configured VLANs. The native VLAN specifies the VLAN on which the port forwards untagged packets it receives.

```
(Switch) #config
(Switch) (Config)#interface 0/8
(Switch) (Interface 0/8)#switchport mode trunk
(Switch) (Interface 0/8)#switchport trunk allowed vlan add 10,20,30
(Switch) (Interface 0/8)#switchport trunk native vlan 100
```

```
(Switch) (Interface 0/8)#exit
```

The General mode port can then be configured as a tagged or untagged member of any VLAN, as shown in “VLAN Configuration Example”.

3.4. Port-channels – Operation and Configuration

Port-channel allows one or more full-duplex (FDX) Ethernet links of the same speed to be aggregated together to form a Port-channel. This allows the switch to treat the Port-channel as if it is a single link. The primary purpose of Port-channels is to increase the overall bandwidth between two switches. This is accomplished by effectively aggregating multiple ports together that act as a single, logical connection between the two switches. Port-channels also provide redundancy. If a link fails, traffic is automatically redistributed across the remaining links.

The switch supports industry-standard Port-channels that adhere to the IEEE 802.3ad specification. Both static and dynamic Port-channels are supported. Each Port-channel can have a maximum of 32 ports as members (as long as the switch can support it). You can configure Port-channels until all switch ports are assigned to a Port-channel.

The following figure shows an example of a switch in the wiring closet connected to a switch in the data center by a Port-channel that consists of four physical 10 Gbps links. The Port-channel provides full-duplex bandwidth of 40 Gbps between the two switches.

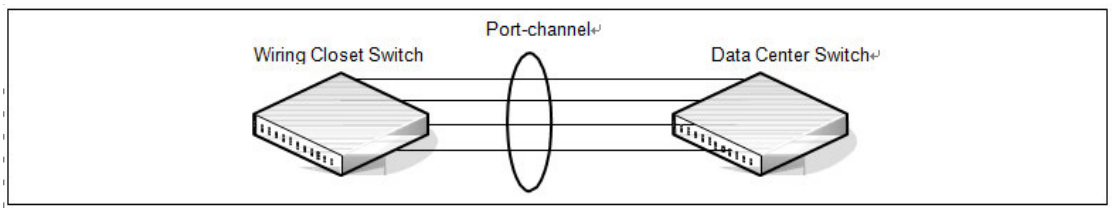


Figure 3-4: Port-channel Configuration

3.4.1. Static and Dynamic Port-channel

Port-channel can be configured as either dynamic or static. Dynamic configuration is supported using the IEEE 802.3ad standard, which is known as Link Aggregation Control Protocol (LACP). Static configuration is used when connecting the switch to an external Gigabit Ethernet switch that does not support LACP.

One advantage of LACP is that the protocol enables the switch to confirm that the external switch is also configured for Port-channel. When using static configuration, a cabling or configuration mistake involving the local switch or the external switch could go undetected and thus cause undesirable network behavior. Both static and dynamic Port-channels (via LACP) can detect physical link failures within the Port-channel and continue forwarding traffic through the other connected links within that same Port-channel. LACP can also detect switch or port failures that do not result in loss of link. This provides a more resilient Port-channel. Best practices suggest using dynamic link aggregation instead of static link aggregation. When a port is added to a Port-channel as a static member, it neither transmits nor receives LACP PDUs.

3.4.2. Port-channel Hashing

The switch supports the configuration of hashing algorithms for each Port-channel interface. The hashing algorithm is used to distribute traffic load among the physical ports of the Port-channel while preserving the per-flow packet order.

The hashing algorithm uses various packet attributes to determine the outgoing physical port. The switch supports the following set of packet attributes to be used for hash computation:

- Source MAC, VLAN, EtherType, and incoming port.
- Destination MAC, VLAN, EtherType, and incoming port.
- Source IP and Source TCP/UDP port numbers.
- Destination IP and Destination TCP/UDP port numbers.
- Source/Destination MAC, VLAN, EtherType, and incoming port.
- Source/Destination IP and Source/Destination TCP/UDP port numbers.
- Enhanced hashing mode

Enhanced hashing mode has following advantages:

- MODULO-N operation based on the number of ports in the Port-channel.
- Packet attributes selection based on the packet type. For L2 packets, Source and Destination MAC address are used for hash computation. For IP packets, Source IP, Destination IP address, TCP/UDP ports are used.
- Non-Unicast traffic and Unicast traffic is hashed using a common hash algorithm.
- Excellent load balancing performance.

3.4.2.1. Resilient Hashing

Resilient Hashing (RH) supports an extra level of indirection between the hash value and the selected output port for a layer-2 Port-channel or a layer-3 ECMP route. In a typical non-RH configuration, the output port can change for all flows when the number of ports changes, even if the flow was on a port that was not affected. This can cause degraded performance due to frame reordering. With RH, the hash value is used to index into a table of ports. If a port goes down, then only the entries that use that port are rewritten. Other ports are left untouched and, therefore, do not suffer degraded performance.

Resilient hashing is globally enabled on switch ports by default. It can be globally enabled (or disabled) in Global Config mode using the **(no) port-channel resilient-hashing** command for Port-channels or the **(no) ip resilient-hashing** command for ECMP routes. The new setting takes effect after a system reboot.

3.4.2.2. Hash Prediction with ECMP and Port-channel

The Hash Prediction feature provides a utility to predict how packets will be forwarded over a Port-channel or to the next- hop device when Equal-Cost Multipath (ECMP) is the destination. Given the Port-channel method, ingress physical port, and values of various packet fields, the utility predicts an egress physical port for the packet.

An ECMP group is identified by the IP address of one of its members. By entering the IP address in the form

<prefix/prefix-length>, the utility predicts the packet's physical egress port based on the destination ECMP group. To predict the an egress physical port when the egress objects are VLAN routing interfaces with Port-channel or port interfaces as members of the VLANs, the utility requires the PVID to be configured on the interfaces and the next hops to be fully installed in hardware.

If an ECMP group is comprised of VLAN routing interfaces and each VLAN has a Port-channel that contains multiple ports, the utility requires the PVID to be configured on the Port-channels. In this configuration, the utility first predicts which VLAN routing interface the packet is forwarded to and finds the Port-channel by matching the VLAN ID of the VLAN routing interface to the PVID of the Port-channel. Then, it predicts which physical port in the Port-channel the packet is forwarded to.

To make correct prediction when Port-channels are used as egress interfaces, the utility requires the enhanced hashing mode to be set on the Port-channels.

Hash prediction is supported for unicast packets only.

3.4.3. Port-channel Interface Overview

The **show interface port-channel brief** command provides summary information about all Port-channels available on the system. In the following output, Port-channel 3/1 has been configured as a dynamic Port-channel with five member ports. No other Port-channels have been configured.

```
(M4500-48XF8C) #show interface port-channel brief


 Channel  Port-Channel Name Min Link State Trap     Type    Mbr Ports  Active Ports

   ID                                    Flag

--------- ----------------- --- ---------- -------- ------- ---------  ------------

1         ch1               1   Down       Disabled Static

2         ch2               1   Down       Disabled Static

3         ch3               1   Down       Disabled Static

4         ch4               1   Down       Disabled Static

5         ch5               1   Down       Disabled Static

6         ch6               1   Down       Disabled Static

7         ch7               1   Down       Disabled Static
```

```
8          ch8               1    Down       Disabled Static

9          ch9               1    Down       Disabled Static

10         ch10              1    Down       Disabled Static

11         ch11              1    Down       Disabled Static

12         ch12              1    Down       Disabled Static

13         ch13              1    Down       Disabled Static

14         ch14              1    Down       Disabled Static

15         ch15              1    Down       Disabled Static

16         ch16              1    Down       Disabled Static

17         ch17              1    Down       Disabled Static

18         ch18              1    Down       Disabled Static

19         ch19              1    Down       Disabled Static
```

3.4.4. Port-channel Interaction with Other Features

From a system perspective, a Port-channel is treated just as a physical port, with the same configuration parameters for administrative enable/disable, spanning tree port priority, path cost as may be for any other physical port.

3.4.4.1. VLAN

When members are added to a Port-channel, they are removed from all existing VLAN membership. When members are removed from a Port-channel they are added back to the VLANs that they were previously members of as per the configuration file. Note that a port's VLAN membership can still be configured when it's a member of a Port-channel. However this configuration is only actually applied when the port leaves the Port-channel.

The Port-channel interface can be a member of a VLAN complying with IEEE 802.1Q.

3.4.4.2. STP

Spanning tree does not maintain state for members of a Port-channel, but the Spanning Tree does maintain state for the Port-channel interface. As far as STP is concerned, members of a Port-channel do not exist. (Internally, the STP state of the Port-channel interface is replicated for the member links.)

When members are deleted from a Port-channel they become normal links, and spanning tree maintains their state information.

3.4.4.3. Statistics

Statistics are maintained for all Port-channel interfaces as they are done for the physical ports, besides statistics maintained for individual members as per the 802.3ad MIB statistics.

3.4.5. Port-channel Configuration Guidelines

Ports to be aggregated must be configured so that they are compatible with the Port-channel feature and with the partner switch to which they connect.

Ports to be added to a Port-channel must meet the following requirements:

- Interface must be a physical Ethernet link.
- Each member of the Port-channel must be running at the same speed and must be in full duplex mode.
- The port cannot be a mirrored port

The following are the interface restrictions

- The configured speed of a Port-channel member cannot be changed.
- An interface can be a member of only one Port-channel.

3.4.5.1. Port-channel Configuration Examples

This section contains the following examples:

- Configuring Dynamic Port-channels
- Configuring Static Port-channels

Note: The examples in this section show the configuration of only one switch. Because Port-channels involve physical links between two switches, the Port-channel settings and member ports must be configured on both switches.

3.4.5.2. Configuration Dynamic Port-channels

The commands in this example show how to configure a dynamic Port-channel on a switch. The Port-channel number is 1 (ch1), and the member ports are 1, 2, 3, 6, and 7.

To configure the switch:

1. Enter interface configuration mode for the ports that are to be configured as Port-channel members.

```
(Switch) #config
(Switch) (Config)#interface range 0/1-0/3,0/6-0/7
```

2. Add the ports to Port-channel 1 with LACP.

```
(Switch) (Interface 0/1-0/3,0/6-0/7)#channel-group 1 mode active
(Switch) (Interface 0/1-0/3,0/6-0/7)#exit
```

3. View information about Port-channel 1.

```
(Switch) #show interface port-channel 1
Port Channel ID.............................. 1
Channel Name................................. ch1
Link State................................... Down
Admin Mode................................... Enabled
Link Trap Mode............................... Enabled
STP Mode..................................... Enabled
Type......................................... Dynamic
Port-channel Min-links....................... 1
Load Balance Option.......................... 3
(Src/Dest MAC, VLAN, EType, incoming port)
LACP Min. Links.............................. 1

Mbr    Device/       Port      Port
Ports  Timeout       Speed     Active
------ ------------- --------- -------
0/1    actor/long    10G Full  False
       partner/long
0/2    actor/long    10G Full  False
       partner/long
0/3    actor/long    10G Full  False
       partner/long
0/6    actor/long    10G Full  False
       partner/long
0/7    actor/long    10G Full  False
       partner/long
```

3.4.5.3. Configuration Static Port-channels

The commands in this example show how to configure a static Port-channel on a switch. The Port-channel number is 3 (ch3), and the member ports are 10, 11, 14, and 17. To configure the switch:

1. Enter interface configuration mode for the ports that are to be configured as Port-channel members.

```
(Switch) (Config)#interface range 0/10-0/12,0/14,0/17
```

2. Add the ports to Port-channel 3 without LACP.

```
(Switch) (Interface 0/10-0/12,0/14,0/17)#channel-group 3 mode on
(Switch) (Interface 0/10-0/12,0/14,0/17)#exit
(Switch) (Config)#exit
```

3. View information about Port-channel 3.

```
(Switch) #show interface port-channel 3
Port Channel ID..............................  3
Channel Name.................................  ch3
Link State...................................  Down
Admin Mode...................................  Enabled
Link Trap Mode...............................  Enabled
STP Mode.....................................  Enabled
Type.........................................  Static
Port-channel Min-links.......................  1
Load Balance Option..........................  3
(Src/Dest MAC, VLAN, EType, incoming port)
LACP Min. Links..............................  1

Mbr    Device/       Port      Port
Ports  Timeout       Speed     Active
------ ------------- --------- -------
0/10   actor/long    10G Full  False
       partner/long
0/11   actor/long    10G Full  False
       partner/long
0/12   actor/long    10G Full  False
       partner/long
0/14   actor/long    10G Full  False
       partner/long
0/17   actor/long    10G Full  False
       partner/long
```

3.5. LACP Fallback Configuration

3.5.1. Configuring Dynamic Port-channels

The commands in this example show how to configure a dynamic Port-channel on a switch. The Port-channel number is 1 (ch1), and the member ports are 1, 2, 3, 6, and 7.

To configure the switch:

1. Enter interface configuration mode for the ports that are to be configured as Port-channel members.

```
(Switch) #config
(Switch) (Config)#interface range 0/1-0/3,0/6-0/7
```

2. Add the ports to Port-channel 1 with LACP.

```
(Switch) (Interface 0/1-0/3,0/6-0/7)#channel-group 1 mode active
(Switch) (Interface 0/1-0/3,0/6-0/7)#exit
```

3. View information about Port-channel 1.

```
(Switch) #show interface port-channel 1
Port Channel ID..............................  1
Channel Name.................................  ch1
Link State...................................  Down
Admin Mode...................................  Enabled
Link Trap Mode...............................  Enabled
STP Mode.....................................  Enabled
Type.........................................  Dynamic
Port-channel Min-links.......................  1
Load Balance Option..........................  3
(Src/Dest MAC, VLAN, EType, incoming port)
LACP Min. Links..............................  1

Mbr    Device/       Port      Port
Ports  Timeout       Speed     Active
------ ------------- --------- -------
0/1    actor/long    10G Full  False
       partner/long
0/2    actor/long    10G Full  False
       partner/long
0/3    actor/long    10G Full  False
       partner/long
```

```
0/6    actor/long    10G Full  False
       partner/long
0/7    actor/long    10G Full  False
       partner/long
```

4. (Optional) Enable LACP Fallback feature which enabled the switch keep one LACP member port link up even if LACP port doesn't receive the LACP message from the other side.

```
(Switch) #(if-port-channel ch1)#lacp fallback
```

5. (Optional) Verify LACP Fallback feature of Port-channel 1.

```
(Switch) #show interface port-channel 1


Port Channel ID.............................. 1
Channel Name................................. ch1
Link State................................... Down
Admin Mode................................... Enabled
Link Trap Mode............................... Disabled
STP Mode..................................... Enabled
Type......................................... Dynamic
Port-channel Min-links....................... 1
Load Balance Option.......................... 3
(Src/Dest MAC, VLAN, EType, incoming port)
LACP Fallback Mode........................... Enabled
LACP Fallback Timeout........................ 5
LACP Min. Links.............................. 1

Mbr    Device/       Port      Port    Fallback
Ports  Timeout       Speed     Active
------ ------------- --------- ------- ---------
0/1    actor/long    10G Full  False   None
       partner/long
0/2    actor/long    10G Full  False   None
```

3.5.2. Configuring Static Port-channels

The commands in this example show how to configure a static Port-channel on a switch. The Port-channel number is 3 (ch3), and the member ports are 10, 11, 14, and 17. To configure the switch:

1. Enter interface configuration mode for the ports that are to be configured as Port-channel members.

```
(Switch) (Config)#interface range 0/10-0/12,0/14,0/17
```

2. Add the ports to Port-channel 3 without LACP.

```
(Switch) (Interface 0/10-0/12,0/14,0/17)#channel-group 3 mode on
(Switch) (Interface 0/10-0/12,0/14,0/17)#exit
(Switch) (Config)#exit
```

3. View information about Port-channel 3.

```
(Switch) #show interface port-channel 3
Port Channel ID..............................  3
Channel Name.................................  ch3
Link State...................................  Down
Admin Mode...................................  Enabled
Link Trap Mode...............................  Enabled
STP Mode.....................................  Enabled
Type.........................................  Static
Port-channel Min-links.......................  1
Load Balance Option..........................  3
(Src/Dest MAC, VLAN, EType, incoming port)
LACP Min. Links..............................  1
Mbr    Device/       Port      Port
Ports  Timeout       Speed     Active
------ ------------- --------- -------
0/10   actor/long    10G Full  False
       partner/long
0/11   actor/long    10G Full  False
       partner/long
0/12   actor/long    10G Full  False
       partner/long
0/14   actor/long    10G Full  False
       partner/long
0/17   actor/long    10G Full  False
       partner/long
```

3.6. MLAG – Operation and Configuration

3.6.1. Overview

In a typical layer-2 network, the Spanning Tree Protocol (STP) is deployed to avoid packet storms due to loops in the network. To perform this function, STP sets ports into either a forwarding state or a blocking state.

Ports in the blocking state do not carry traffic. In the case of a topology change, STP reconverges to a new loop-free network and updates the port states. STP is relatively successful mitigating packet storms in the network, but redundant links in the network are blocked from carrying traffic by the spanning tree protocol.

In some network deployments, redundant links between two switches are bundled together in a Port-channel and appear as a single link in the spanning tree topology. The advantage is that all Port-channel member links can be in the forwarding state and a link failure can be recovered in milliseconds. This allows the bandwidth on the redundant links to be utilized. However, Port-channels are limited to connecting multiple links between two partner switches, which leaves the switch as a single point of failure in the topology.

The MLAG extends the Port-channel bandwidth advantage across multiple switches connected to a Port-channel partner device. The Port-channel partner device is oblivious to the fact that it is connected over a Port-channel to two peer switches; instead, the two switches appear as a single switch to the partner with a single MAC address. All links can carry data traffic across a physically diverse topology and in the case of a link or switch failure, traffic can continue to flow with minimal disruption.

3.6.2. Deployment Scenarios

MLAG is intended to support higher bandwidth utilization in scenarios where a redundant layer-2 network is desired. In such scenarios the effects of STP on link utilization are profound. Large percentages of links do not carry data because they are blocked and only a single path through the network carries traffic.

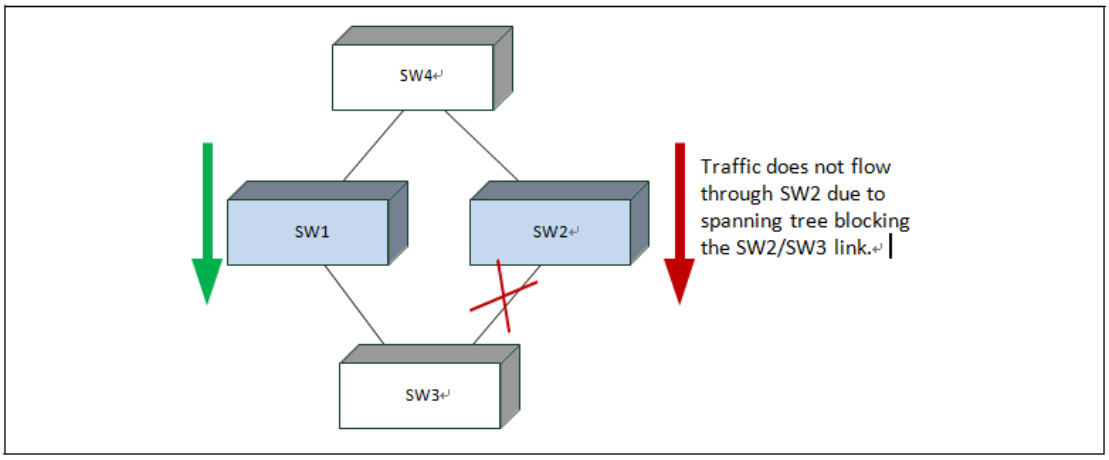


Figure 3-5: STP Blocking

MLAG reduces some of the bandwidth shortcomings of STP in a layer-2 network. It provides a reduced convergence period when a port-channel link goes down and provides more bandwidth because all links can forward traffic. In the figure below, if SW1 and SW2 form a MLAG with SW3 and SW4, none of the links are blocked, which means traffic can flow over both links from SW4 through to SW1 and SW2 over both links from SW1 and SW2 to SW3.

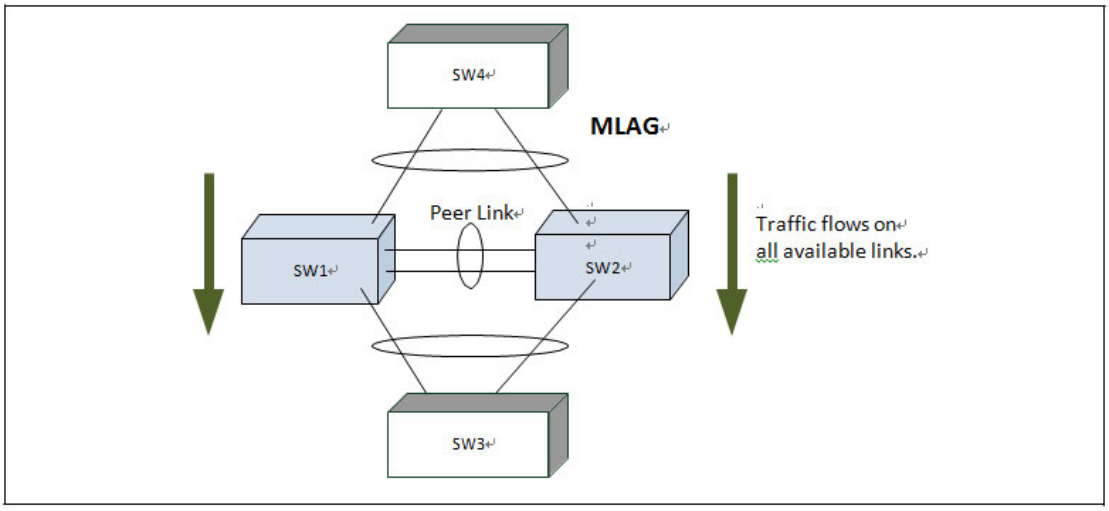


Figure 3-6: MLAG in a Layer-2 Network

3.6.2.1. Definitions

Refer to the following figure for the definitions that follow the figure.

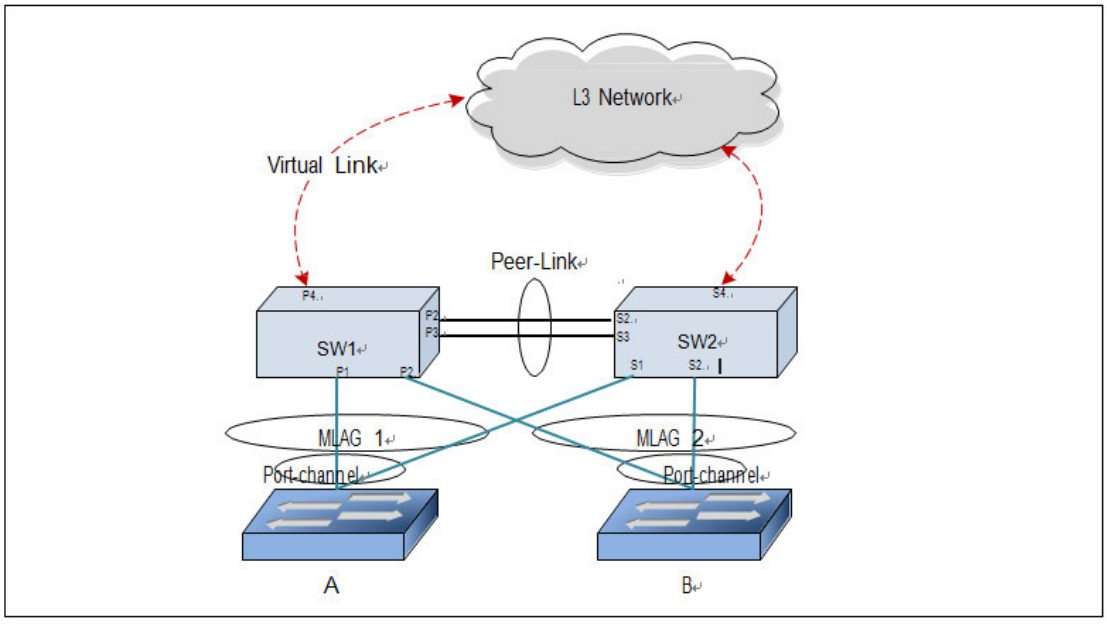


Figure 3-7: MLAG Components

MLAG switches: MLAG-aware switches running switch firmware. No more than two MLAG aware switches can pair to form one end of the Port-channel. In the previous figure, SW1 and SW2 are MLAG peer switches. These two switches form a single logical end point for the MLAG from the perspective of switch A.

MLAG interfaces: MLAG functionality is a property of Port-channels. Port-channels configured as MLAGs are called MLAG interfaces. Administrators can configure multiple instances of MLAG interfaces on the peer

MLAG switches. Port-channel limitations and capabilities such as min-links and maximum number of ports supported per Port-channel also apply to MLAG interfaces.

MLAG member ports: Ports on the peer MLAG switches that are part of the MLAG interface (P1 on SW1 and S1 on SW2).

MLAG peer-link: A link between the two MLAG peer switches (ports P2, P3, S2, S3). Only one peer-link can be configured per device. The peer-link is crucial for the operation of the MLAG component. A Port-channel must be configured as the peer-link. All VLANs configured on MLAG interfaces must be configured on the peer-link as well.

MLAG Dual Control Plane Detection link: A virtual link that is used to advertise the Dual Control Plane Detection Protocol (DCPDP) packets between the two MLAG switches (ports P4, S4). DCPDP is optional but should be used with caution. The protocol is used as a secondary means of detecting the presence of the peer switch in the network. The DCPDP protocol must not be configured on MLAG interfaces.

3.6.2.2. Configuration Consistency

MLAG is operational only if the MLAG domain ID, MLAG system MAC address, and MLAG system priority are the same on both the MLAG peer switches.

Note: Configuring a MLAG domain ID is mandatory; the MLAG system MAC address and MLAG system priority are optional (these values are auto generated if not configured)

You must ensure that the neighboring devices connected to MLAG switches perceive the two switches as a single spanning tree and Link Aggregation Control Protocol (LACP) entity. To achieve this end, the following configuration settings must be identical for MLAG links on the MLAG peer switches:

1. Port-channel

- Hashing mode
- Minimum links
- Static/dynamic Port-channel
- LACP parameters

  – Actor parameters

  – Admin key

  – Collector max-delay

  – Partner parameters

2. STP

The default STP mode is MSTP. The following STP configuration parameters must be the identical on both MLAG peers.

- Spanning-tree version (RSTP)
- Bpdufilter
- Bpduflood
- Auto-edge
- TCN-guard
- Cost
- Edgeport
- STP Version
- Root guard
- Loop guard

3. Port-channel interface

The following Port-channel attributes must be identical for MLAG Port-channels:

- Port-channel mode
- Link speed
- Duplex mode
- MTU
- Bandwidth
- VLAN configuration

You must also ensure that the following are identical before enabling MLAG:

- FDB entry aging timers
- Static MAC entries.
- ACL configuration

4. Interface Configuration

- PFC configuration
- CoS queue assignments

5. VLAN configuration

   - MLAG VLANs must span the MLAG topology and be configured on both MLAG peers. This means that every MLAG VLAN must connect to two partner Port-channels.
   - VLAN termination of a MLAG VLAN on a MLAG peer is not supported.

6. Switch firmware versions

Except during firmware upgrade, the peer switch firmware versions must be identical, as subtle differences between versions may cause instability.

You must ensure that the above configuration items are configured identically on the MLAG interfaces on both of the MLAG peers before enabling the MLAG feature. If the configuration settings are not in sync, the MLAG behavior is undefined. Once the above configuration is in place and consistent, the two switches will form a MLAG that operates in the desired manner. The MLAG may form even if the configuration is not consistent, however, it may not operate consistently in all situations.

3.6.3. MLAG Fast Failover

If a switch does not support MLAG fast failover, when the primary switch fails, the secondary switch restarts the LACP protocol on its MLAG member ports. STP is also restarted on the secondary device's MLAG member ports. Until the LACP and STP reconverges, the partner device is disconnected from the MLAG domain.

With fast failover support, neither LACP reconvergence nor STP reconvergence occurs, and minimal traffic loss is observed when primary device fails. During the failover, traffic that is being forwarded using the links connected to primary device will failover to links connected to the secondary device. The traffic disruption is limited to the time required for the partner devices dual-attached to the MLAG domain to detect the link down (links connected to primary device) and redistribute the traffic using the links connected to the secondary device.

3.6.4. MLAG Configuration

Refer to the following figure for a visual overview of the MLAG configuration steps that follow the figure.

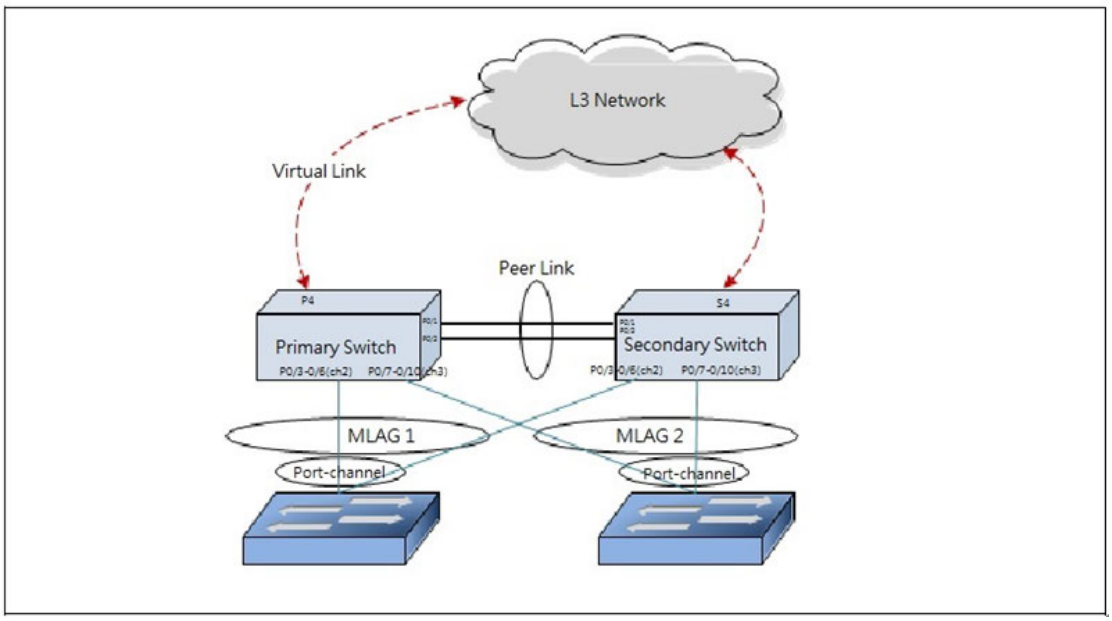


Figure 3-8: MLAG Configuration Diagram

To configure MLAG:

1. Enter VLAN data base mode and create the MLAG VLANs.

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 1-100
```

2. Enable the MLAG feature.

```
(Switch) #config
(Switch) (Config)#mlag
```

3. Create the MLAG domain ID. The domain ID configured on both the MLAG peer switches should be same. In a two-tier MLAG topology, each pair should have different domain ID.

```
(Switch) (Config)#mlag domain 1
```

4. Configure the MLAG system MAC address and/or MLAG system priority (optional).

```
(Switch) (Config)#mlag system-mac C4:54:44:01:01:01
```

5. Enable the keepalive protocol.

```
(Switch) (Config)#mlag peer-keepalive enable
```

6. Configure the MLAG role priority (optional).

```
(Switch) (Config)#mlag role priority 10
```

7. Create Port-channel 1.

```
(Switch) (Config)#interface port-channel 1
(Switch) (if-port-channel ch1)#description "MLAG-Peer-Link"
```

8. Allow the Port-channel to participate in all VLANs and accept and send tagged frames only. This is similar to configuring a port in trunk mode.

```
(Switch) (if-port-channel ch1)#switchport allowed vlan add tagged 1-99
(Switch) (if-port-channel ch1)#switchport acceptable-frame-types tagged
(Switch) (if-port-channel ch1)#mlag peer-link
(Switch) (if-port-channel ch1)#exit
```

9. Create the peer link.

```
(Switch) (Config)#interface range 0/1-0/2
(Switch) (Interface 0/1-0/2)#channel-group 1 mode active
(Switch) (Interface 0/1-0/2)#description "MLAG-Peer-Link"
```

10. Enable UDLD (if required).

```
(Switch) (Interface 0/1-0/2)#udld enable
(Switch) (Interface 0/1-0/2)#udld port aggressive
(Switch) (Interface 0/1-0/2)#exit
```

11. Configure Dual Control Plane detection Protocol Configuration (if required):

a. Configure the peer-switch IP address (the destination IP address, serviceport is recommended). This command configures the IP address of the peer MLAG switch. This configuration is used by the dual control plane detection protocol (DCPDP) on the MLAG switches. The UDP port on which the MLAG switch listens to the DCPDP messages can also be configured with this command. The configurable range for the UDP port 1 to 65535 (Default is 60000).

```
(Switch) (Config)#serviceport protocol none
(Switch) (Config)#serviceport ip 192.168.0.2 255.255.255.0 192.168.0.254
```

b. Configure the keepalive source and destination IP address.

```
(Switch) #config
(Switch) (Config)#mlag peer-keepalive destination 192.168.0.1 source 192.168.0.2
```

12. Configure a Port-channel as MLAG interface. The configurable range for the MLAG ID is 1 to 63.

```
(Switch) (Config)#interface range 0/3-0/6
(Switch) (Interface 0/3-0/6)#channel-group 2 mode active
(Switch) (Interface 0/3-0/6)#exit
(Switch) (Config)#interface range 0/7-0/10
(Switch) (Interface 0/7-0/10)#channel-group 3 mode active
(Switch) (Interface 0/7-0/10)#exit
(Switch) (Config)#interface port-channel 2
(Switch) (if-port-channel ch2)#switchport allowed vlan add tagged 1-99
(Switch) (if-port-channel ch2)#switchport acceptable-frame-types tagged
(Switch) (if-port-channel ch2)#mlag 1
```

```
(Switch) (if-port-channel ch2)#exit
(Switch) (Config)#interface lag 3
(Switch) (if-port-channel ch3)#switchport allowed vlan add tagged 1-99
(Switch) (if-port-channel ch3)#switchport acceptable-frame-types tagged
(Switch) (if-port-channel ch3)#mlag 2
(Switch) (if-port-channel ch3)#exit
```

13. MLAG can support to work with RSTP to provide the loop prevention mechanism. To prevent the user error connection and lead the network environment crash by broadcast storm.

```
(Switch) (Config)#spanning-tree mode rstp
(Switch) (Config)#spanning-tree
```

14. IGMP snooping support is provided. If the network environment need multicast traffic, MLAG can enable IGMP snooping, and IGMP snooping and MLAG can cooperate.

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 40
(Switch) (Vlan)#set igmp 40
(Switch) (Vlan)#exit
(Switch) (Config)#ip igmp snooping
```

You must ensure that the port channel configurations on both devices are in sync before enabling MLAG. After the MLAG interfaces are enabled, the MLAG interfaces are operationally shut down. The MLAG component exchanges information regarding the port members that constitute the Port-channel on each device. Once this information is populated on both devices, the MLAG interfaces are operationally up and traffic forwarding on MLAG interfaces is allowed. Port-channels must be configured on both devices as MLAG interfaces for the MLAG interface to be enabled. Also, the port-channel-number: MLAG-Id pair must be the same on both the primary and secondary devices.

Member ports can be added or removed from the MLAG interface. If a port is added as a port member to a MLAG interface, the Primary allows the port member if the maximum criteria is satisfied. When a port member is removed from the MLAG interface, the Primary decides if the minimum criteria is satisfied. If it is not, it will shut down the MLAG interface on both the devices. Shutting down the MLAG interface on the Secondary is not allowed. The MLAG interface can only be shut down on the Primary.

FDB entries learned on MLAG interfaces are synced between the two devices. In the case where all MLAG member ports are UP, data traffic does not traverse the peer link.

3.7. Unidirectional Link Detection (UDLD)

The UDLD feature detects unidirectional links on physical ports. UDLD must be enabled on the both sides of the link in order to detect a unidirectional link. The UDLD protocol operates by exchanging packets containing information about neighboring devices.

The purpose of UDLD feature is to detect and avoid unidirectional links. A unidirectional link is a forwarding anomaly in a Layer 2 communication channel in which a bidirectional link stops passing traffic in one direction.

3.7.1. UDLD Modes

The UDLD supports two modes: normal and aggressive.

In normal mode, a port's state is classified as *undetermined* if an anomaly exists. An anomaly might be the absence of its own information in received UDLD messages or the failure to receive UDLD messages. An *undetermined* state has no effect on the operation of the port. The port is not disabled and continues operating. When operating in UDLD normal mode, a port will be put into a disabled state (D-Disable) only in the following situations:

- The UDLD PDU received from a partner does not have its own details (echo).
- When there is a loopback, and information sent out on a port is received back exactly as it was sent.

When operating in UDLD aggressive mode, a port is put into a disabled state for the same reasons that it occurs in normal mode. Additionally, a port in UDLD aggressive mode can be disabled if the port does not receive any UDLD echo packets even *after* bidirectional connection was established. If a bidirectional link is established, and packets suddenly stop coming from partner device, the UDLD aggressive-mode port assumes that link has become unidirectional.

3.7.2. UDLD and Port-channel Interfaces

UDLD is supported on individual physical ports that are members of a Port-channel. If any of the aggregated links becomes unidirectional, UDLD detects it and disables the individual link, but not the entire Port-channel. This improves the fault tolerance of the Port-channel.

3.7.3. Configuring UDLD

A network administrator decides to use the UDLD feature while building a loop-free topology with the use of STP. You must configure the ports on both side of the link to use UDLD in aggressive mode to ensure that ports with unidirectional links will be shut down, and no loops will be introduced into topology. This example shows the steps to configure UDLD on Switch 1 only. The same configuration must be performed on all ports that form partner links with the ports on Switch 1.

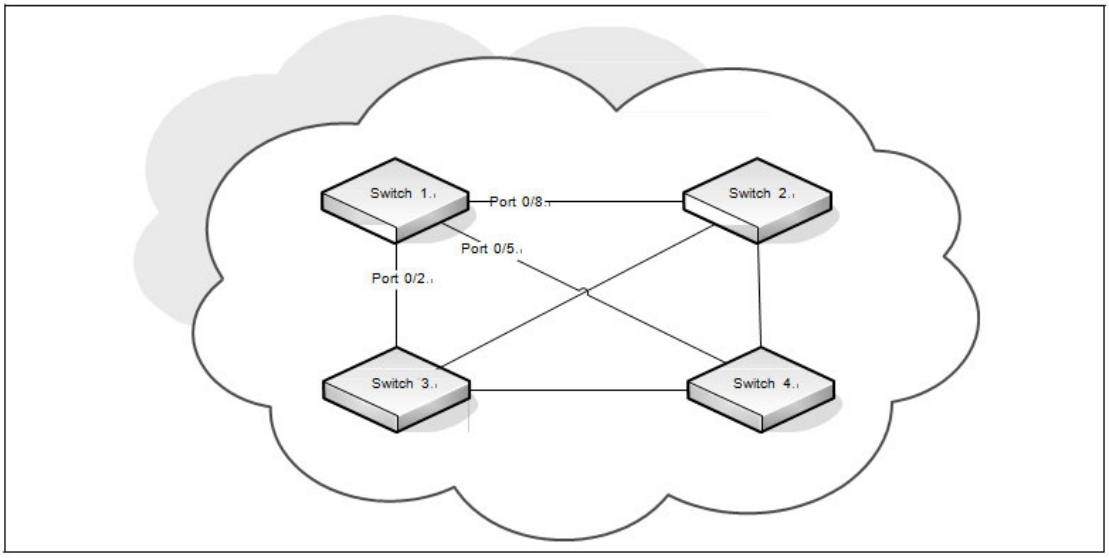


Figure 3-9: UDLD Configuration Example

To configure the ports on Switch 1:

1. Globally enable UDLD on the switch.

```
(Switch) #configure
(Switch) (Config)#udld enable
```

2. Enter interface configuration mode for the ports that are connected to other switches and enable UDLD on the ports.

```
(Switch) (Config)#interface range 0/2,0/5,0/8
(Switch) (Interface 0/2,0/5,0/8)#udld enable
```

3. Configure the UDLD mode on the ports to be aggressive.

```
(Switch) (Interface 0/2,0/5,0/8)#udld port aggressive
(Switch) (Interface 0/2,0/5,0/8)#exit
(Switch) (Config)#exit
```

1.After configuring UDLD on Switch 2, Switch, 3, and Switch 4, view the UDLD status for the ports.

```
(Switch) #show udld all


Port     Admin Mode     UDLD Mode      UDLD Status
-----    ----------     -----------    --------------
0/1      Disabled       Normal         Not Applicable
0/2      Enabled        Aggressive     Bidirectional
0/3      Disabled       Normal         Not Applicable
0/4      Disabled       Normal         Not Applicable
0/5      Enabled        Aggressive     Bidirectional
0/6      Disabled       Normal         Not Applicable
0/7      Disabled       Normal         Not Applicable
0/8      Enabled        Aggressive     Bidirectional
```

```
 0/9     Disabled      Normal        Not Applicable

--More-- or (q)uit
```

Note: If a port has become disabled by the UDLD feature and you want to re-enable the port, use the `udld reset` command in Privileged EXEC mode.

3.8. Port Mirroring

Port mirroring is used to monitor the network traffic that a port sends and receives. The Port Mirroring feature creates a copy of the traffic that the source port handles and sends it to a destination port. The source port is the port that is being monitored. The destination port is monitoring the source port. The destination port is where you would connect a network protocol analyzer to learn more about the traffic that is handled by the source port.

A port monitoring session includes one or more source ports that mirror traffic to a single destination port. The switch supports a single port monitoring session. Port-channels cannot be used as the source or destination ports.

For each source port, you can specify whether to mirror ingress traffic (traffic the port receives, or RX), egress traffic (traffic the port sends, or TX), or both ingress and egress traffic.

The packet that is copied to the destination port is in the same format as the original packet on the wire. This means that if the mirror is copying a received packet, the copied packet is VLAN tagged or untagged as it was received on the source port. If the mirror is copying a transmitted packet, the copied packet is VLAN tagged or untagged as it is being transmitted on the source port.

After you configure the port mirroring session, you can enable or disable the administrative mode of the session to start or stop the probe port from receiving mirrored traffic.

3.8.1. Configuring Port Mirroring

In this example, traffic from ports 1 and 4 is mirrored to probe port 10.

1. Configure the source ports. Traffic received and transmitted on by these ports will be mirrored.

```
(Switch) #configure
(Switch) (Config)#port-monitor session 1 source interface 0/1
```

2. Configure the destination (probe) port.

```
(Switch) (Config)#port-monitor session 1 destination interface 0/10
```

3. Enable port mirroring on the switch.

```
(Switch) (Config)#port-monitor session 1 mode
(Switch) (Config)#exit
```

4. View summary information about the port mirroring configuration.

```
(M4500-48XF8C) (Config)#show port-monitor session 1

Session Admin   Probe  Src  Mirrored                Ref.   Src   Dst   Type  IP      MAC
ID      Mode    Port   VLAN Port                    Port   RVLAN RVLAN       ACL     ACL
------- ------- ------ ---- ----------------------- ------ ----- ----- ----- ------- -------
1       Enable  0/10        0/1                                        Rx,Tx
```

3.8.2. Configuring RSPAN

This example mirrors traffic from port 6 on a source switch (SW1) to a probe port on a remote switch (port 12 on SW3). The mirrored traffic is carried in the RSPAN VLAN and VLAN 100, which traverses an intermediate switch (SW2). The commands in this example show how to configure port mirroring on the source, intermediate, and destination switches.

The following figure provides a visual overview of the RSPAN configuration example.

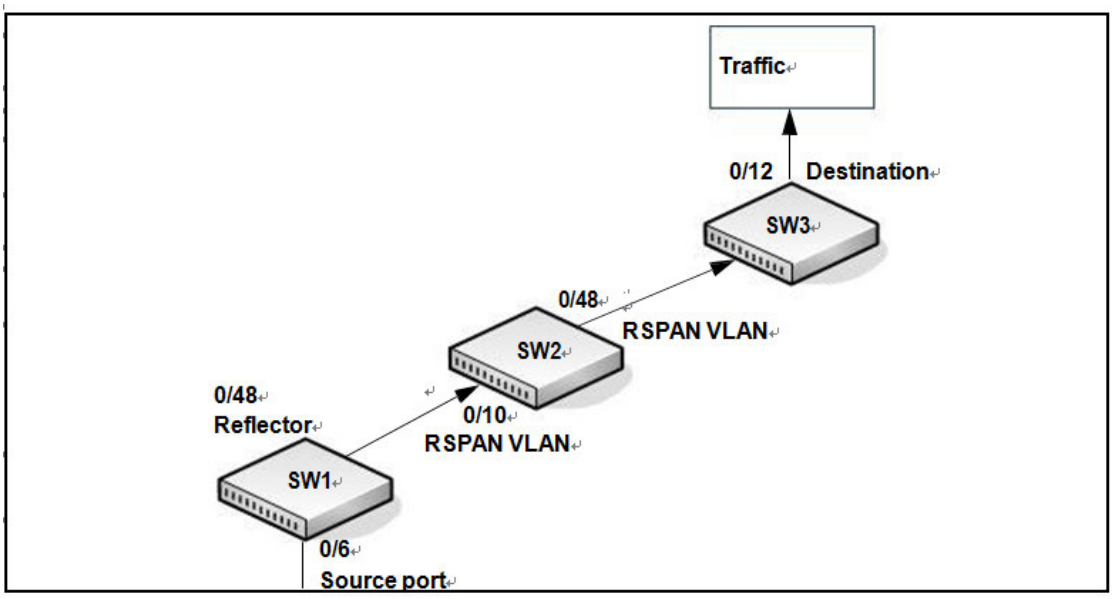


Figure 3-10: RSPAN Configuration Example

3.8.2.1. Configuration on the Source Switch (SW1)

To configure the source switch:

1. Access the VLAN configuration mode and create VLAN 100, which will be the RSPAN VLAN.

```
(Switch) #configure
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 100 (Switch) (Vlan)#exit
```

2. Configure VLAN 100 as the RSPAN VLAN.

```
(Switch) #configure
(Switch) (Config)#vlan 100
(Switch) (Config)(vlan 100)#remote-span
(Switch) (Config)(vlan 100)#exit
```

3. Configure the RSPAN VLAN as the destination port and the reflector port as port 0/48.

```
(Switch) #configure
(Switch) (Config)#port-monitor session 1 destination remote vlan 100 reflector-port 0/48
```

4. Configure the source interface port as port 0/6.

```
(Switch) (Config)#port-monitor session 1 source interface 0/6
```

5. Enable the port mirroring session on the switch.

```
(Switch) (Config)#port-monitor session 1 mode
(Switch) #exit
```

3.8.2.2. Configuration on the Intermediate Switch (SW2)

To configure the intermediate switch:

1. Access the VLAN configuration mode and create VLAN 100.

```
(Switch) #configure
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 100
(Switch) (Vlan)#exit
```

2. Enable RSPAN on vlan 100.

```
(Switch) #configure
(Switch) (Config)#vlan 100
(Switch) (Config)(vlan 100)#remote-span
(Switch) (Config)(vlan 100)#exit
```

3. Configure VLAN participation so the interface is always a member of the VLAN.

```
(Switch) (Config)#interface 0/10
(Switch) (Interface 0/10)#switchport allowed vlan add tagged 100
(Switch) (Interface 0/10)#exit
```

4. Configure VLAN participation so the interface is always a member of the VLAN.

```
(Switch) (Config)#interface 0/48
(Switch) (Interface 0/48)#switchport allowed vlan add tagged 100
(Switch) (Interface 0/48)#exit
```

3.8.2.3. Configuration on the Destination Switch (SW3)

3.8.3. VLAN-based Mirroring

In this example, traffic from all ports that are members of VLAN 10 is mirrored to port 0/18. To configure VLAN based mirroring:

1. Access VLAN Config mode and create VLAN 10.

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 10
(Switch) (Vlan)#exit
```

2. Configure the destination interface port as port 0/18.

```
(Switch) #configure
(Switch) (Config)#port-monitor session 1 destination interface 0/18
```

3. Configure VLAN 10 as the source interface for the port mirroring session.

```
(Switch) (Config)#port-monitor session 1 source vlan 10
```

4. Enable the port mirroring session on the switch.

```
(Switch) (Config)#port-monitor session 1 mode
(Switch) (Config)#exit
```

3.8.4. Flow-based Mirroring

In this example, traffic from port 1 is mirrored to port 18 if it matches the criteria defined in the IP ACL or MAC ACL that are associated with the port mirroring session.

To configure flow based mirroring:

1. Create the extended IP access list IPACL.

```
(Switch) #configure
(Switch) (Config)#ip access-list IPACL
(Switch) (Config-ipv4-acl)#permit ip 1.1.1.1 0.0.0.0 any
(Switch) (Config-ipv4-acl)#exit
```

2. Create the mac access list MACL.

```
(Switch) #configure
(Switch) (Config)#mac access-list extended MACL
(Switch) (Config-mac-access-list)#permit 00:00:00:00:00:11 00:00:00:00:00:00 any
(Switch) (Config-mac-access-list)#exit
```

3. Configure the destination port as port 0/18.

```
(Switch) (Config)#port-monitor session 1 destination interface 0/18
```

4. Configure the source port as port 0/2.

```
(Switch) (Config)#port-monitor session 1 source interface 0/2
```

5. Enable the port mirroring session.

```
(Switch) (Config)#port-monitor session 1 mode
```

6. To filter L3 traffic so only flows that match the rules in the IP ACL called IPACL are mirrored to the destination port, add the IPACL ACL.

```
(Switch) (Config)#port-monitor session 1 filter ip access-group IPACL
```

7. To filter L2 traffic so only flows that match the rules in the MAC-based ACL called MACL are mirrored to the destination port, add the MACL ACL.

```
(Switch) (Config)#port-monitor session 1 filter mac access-group MACL
(Switch) (Config)#exit
```

Note: Both IP ACL and MAC ACL cannot be configured for one session at the same time.

3.9. Spanning Tree Protocol

Spanning Tree Protocol (STP) is a layer 2 protocol that provides a tree topology for switches on a bridged LAN. STP allows a network to have redundant paths without the risk of network loops. STP uses the spanning-tree algorithm to provide a single path between end stations on a network.

The switch supports Multiple STPand Rapid STP.

3.9.1. Classic STP, Multiple STP, and Rapid STP

Classic STP provides a single path between end stations, avoiding and eliminating loops. Multiple Spanning Tree Protocol (MSTP) is specified in IEEE 802.1s and supports multiple instances of Spanning Tree to efficiently channel VLAN traffic over different interfaces. Each instance of the Spanning Tree behaves in the manner specified in IEEE 802.1w, Rapid Spanning Tree (RSTP), with slight modifications in the working but not the end effect (chief among the effects, is the rapid transitioning of the port to Forwarding). The difference between the RSTP and the traditional STP (IEEE 802.1D) is the ability to configure and recognize full-duplex connectivity and ports which are connected to end stations, resulting in rapid transitioning of the port to the Forwarding state and the suppression of Topology Change Notifications.

MSTP is compatible to both RSTP and STP. It behaves appropriately to STP and RSTP bridges. A MSTP bridge can be configured to behave entirely as a RSTP bridge or a STP bridge.

3.9.2. STP Operation

The switches (bridges) that participate in the spanning tree elect a switch to be the root bridge for the spanning tree. The root bridge is the switch with the lowest bridge ID, which is computed from the unique identifier of the bridge and its configurable priority number. When two switches have an equal bridge ID value, the switch with the lowest MAC address is the root bridge.

After the root bridge is elected, each switch finds the lowest-cost path to the root bridge. The port that connects the switch to the lowest-cost path is the root port on the switch. The switches in the spanning tree also determine which ports have the lowest-path cost for each segment. These ports are the designated ports. Only the root ports and designated ports are placed in a forwarding state to send and receive traffic. All other ports are put into a blocked state to prevent redundant paths that might cause loops.

To determine the root path costs and maintain topology information, switches that participate in the spanning tree use Bridge Protocol Data Units (BPDUs) to exchange information.

3.9.3. MSTP in the Network

In the following diagram of a small 802.1D bridged network, STP is necessary to create an environment with full connectivity and without loops.

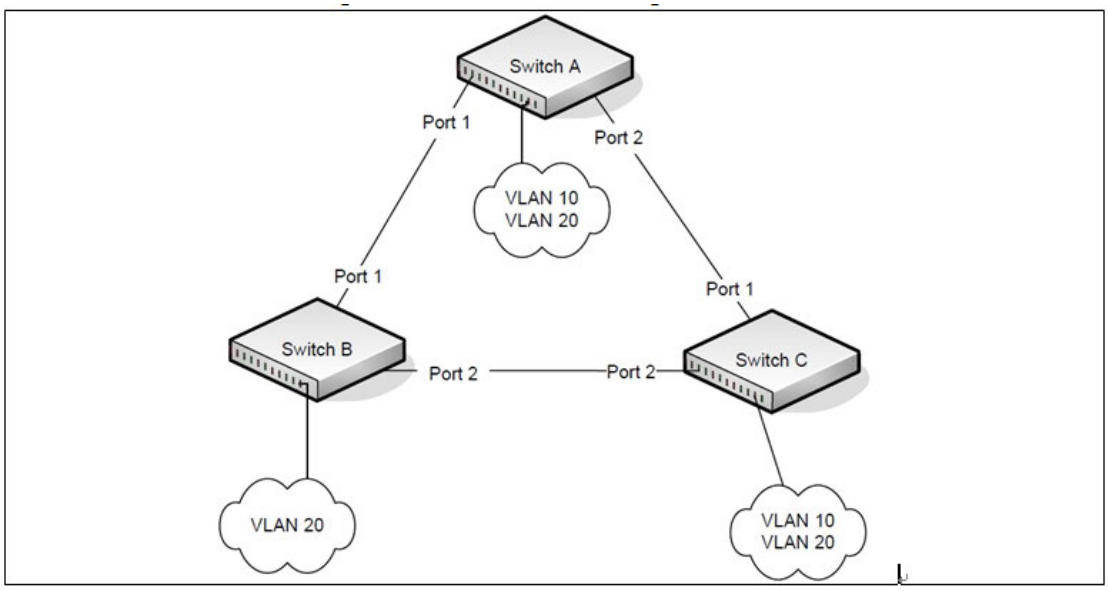


Figure 3-11: STP in a Small Bridged Network

Assume that Switch A is elected to be the Root Bridge, and Port 1 on Switch B and Switch C are calculated to be the root ports for those bridges, Port 2 on Switch B and Switch C would be placed into the Blocking state. This creates a loop-free topology. End stations in VLAN 10 can talk to other devices in VLAN 10, and end stations in VLAN 20 have a single path to communicate with other VLAN 20 devices.

The following figure shows the logical single STP network topology.

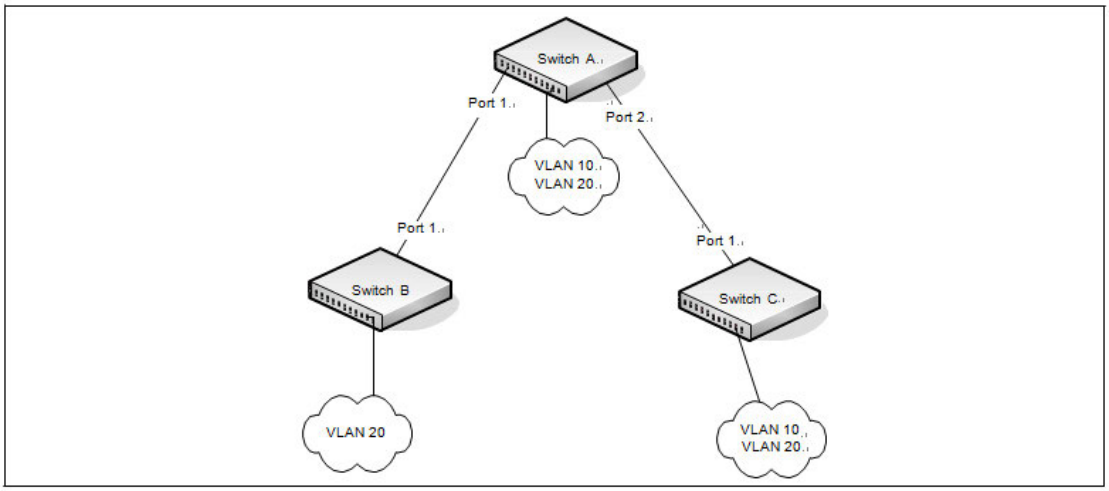


Figure 3-12: Single STP Topology

For VLAN 10 this single STP topology is fine and presents no limitations or inefficiencies. On the other hand, VLAN 20's traffic pattern is inefficient. All frames from Switch B will have to traverse a path through Switch A before arriving at Switch C. If the Port 2 on Switch B and Switch C could be used, these inefficiencies could be eliminated. MSTP does just that, by allowing the configuration of MSTIs based upon a VLAN or groups of VLANs. In this simple case, VLAN 10 could be associated with Multiple Spanning Tree Instance (MSTI)1 and VLAN 20 could be associated with MSTI 2 where Port 1 on both Switch A and Switch B begin discarding and all others forwarding. This simple modification creates an active topology with a better distribution of network traffic and an increase in available bandwidth.

The logical representation of the MSTP environment for these three switches is shown in the following figure.

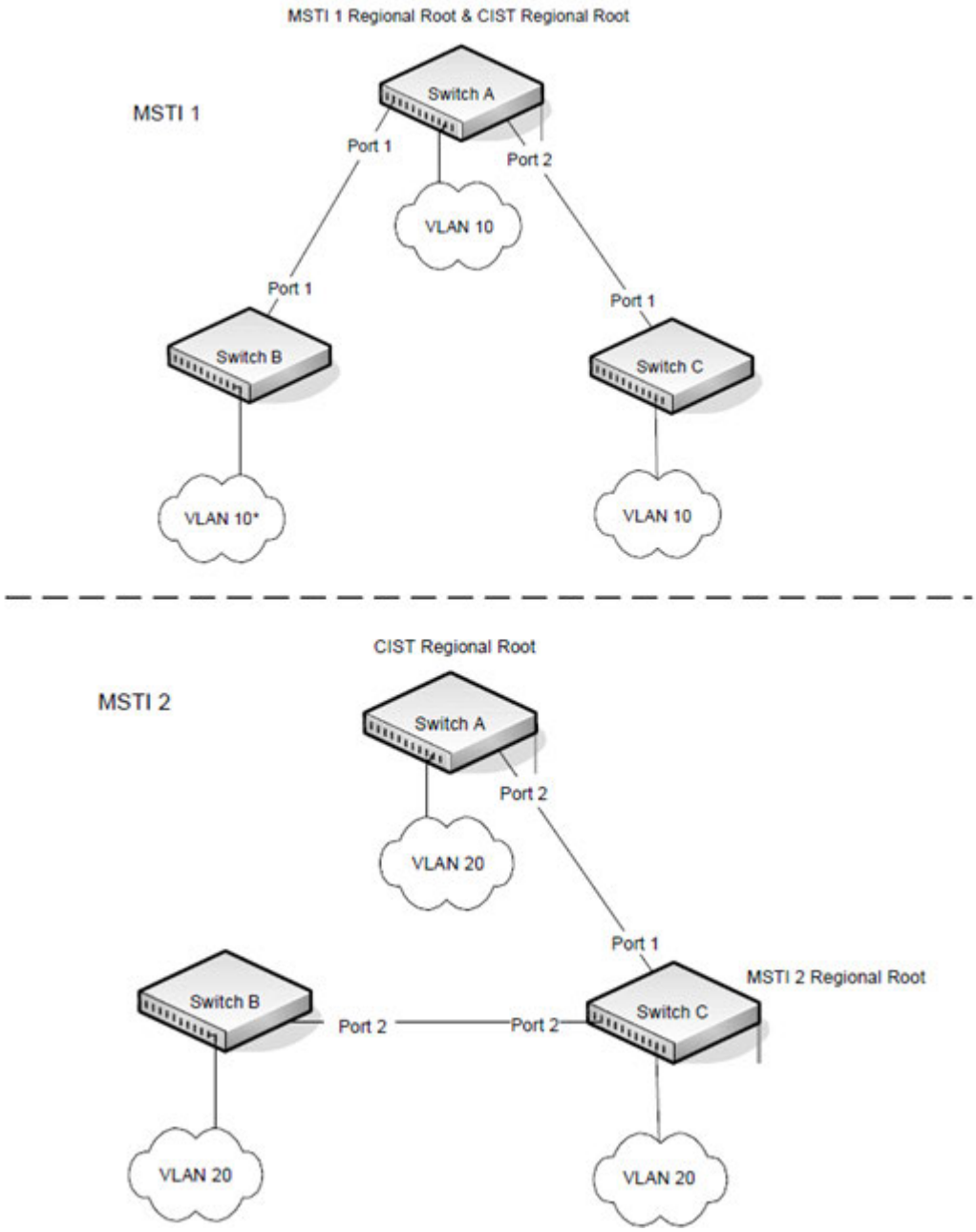


Figure 3-13: Logical MSTP Environment

For MSTP to correctly establish the different MSTIs as above, some additional changes are required. For example, the configuration would have to be the same on each and every bridge. That means that Switch B would have to add VLAN 10 to its list of supported VLANs. This is necessary with MSTP to allow the formation of Regions made up of all switches that exchange the same MST Configuration Identifier. It is within only these MST Regions that multiple instances can exist. It will also allow the election of Regional Root Bridges for each instance. One common and internal spanning tree (CIST) Regional Root for the CIST and an MSTI Regional Root Bridge per instance will enable the possibility of alternate paths through each Region. Above Switch A is elected as both the MSTI 1 Regional Root and the CIST Regional Root Bridge, and after adjusting the Bridge Priority on Switch C in MSTI 2, it would be elected as the MSTI 2 Regional Root.

To further illustrate the full connectivity in an MSTP active topology, the following rules apply:

1. Each Bridge or LAN is in only one Region.

2. Every frame is associated with only one VID.

3. Frames are allocated either to the IST or MSTI within any given Region.

4. The internal spanning tree (IST) and each MSTI provides full and simple connectivity between all LANs and

Bridges in a Region.

5. All Bridges within a Region reach a consistent agreement as to which ports interconnect that Region to a different Region and label those as Boundary Ports.

6. At the Boundary Ports, frames allocated to the CIST or MSTIs are forwarded or not forwarded alike.

7. The CIST provides full and simple connectivity between all LANs and Bridges in the network.

3.9.4. Optional STP Features

The switch supports the following optional STP features:

•BPDU flooding

•Edge Port

•Root guard

•Loop guard

•BPDU protection

3.9.4.1. BPDU Flooding

The BPDU flooding feature determines the behavior of the switch when it receives a BPDU on a port that is disabled for spanning tree. If BPDU flooding is configured, the switch will flood the received BPDU to all the ports on the switch which are similarly disabled for spanning tree.

3.9.4.2. Edge Port

The Edge Port feature reduces the STP convergence time by allowing ports that are connected to end devices (such as a desktop computer, printer, or file server) to transition to the forwarding state without going through the listening and learning states.

3.9.4.3. Root Guard

Enabling root guard on a port ensures that the port does not become a root port or a blocked port. When a switch is elected as the root bridge, all ports are designated ports unless two or more ports of the root bridge are connected together. If the switch receives superior STP BPDUs on a root-guard enabled port, the root guard feature moves this port to a root-inconsistent STP state, which is effectively equal to a listening state. No traffic is forwarded across this port. In this way, the root guard feature enforces the position of the root bridge.

When the STP mode is MSTP, the port may be a designated port in one MSTI and an alternate port in the CIST, and so on. Root guard is a per-port configuration (not a per-port per-instance command), so that all the MSTP instances that this port participates in are not in a root role.

3.9.4.4. Loop Guard

Loop guard protects a network from forwarding loops induced by BPDU packet loss. The reasons for failing to receive packets are numerous, including heavy traffic, software problems, incorrect configuration, and unidirectional link failure. When a non-designated port no longer receives BPDUs, the spanning-tree algorithm considers that this link is loop free and begins transitioning the link from blocking to forwarding. Once in forwarding state, the link may create a loop in the network.

Enabling loop guard prevents such accidental loops. When a port is no longer receiving BPDUs and the max age timer expires, the port is moved to a *loop-inconsistent blocking state*. In the loop-inconsistent blocking state, traffic is not forwarded so the port behaves as if it is in the blocking state. The port will remain in this state until it receives a BPDU. It will then transition through the normal spanning tree states based on the information in the received BPDU.

Note: Loop Guard should be configured only on non-designated ports. These include ports in alternate or backup roles. Root ports and designated ports should not have loop guard enabled so that they can forward traffic

3.9.4.5. BPDU Protection

When the switch is used as an access layer device, most ports function as edge ports that connect to a device such as a desktop computer or file server. The port has a single, direct connection and is configured as an edge port to implement the fast transition to a forwarding state. When the port receives a BPDU packet, the system sets it to non-edge port and recalculates the spanning tree, which causes network topology flapping. In normal cases, these ports do not receive any BPDU packets. However, someone may forge BPDU to maliciously attack the switch and cause network flapping.

BPDU protection can be enabled in RSTP to prevent such attacks. When BPDU protection is enabled, the switch disables an edge port that has received BPDU and notifies the network manager about it.

3.9.5. STP Configuring Examples

3.9.5.1. Configuring MSTP

This example shows how to configure IEEE 802.1s Multiple Spanning Tree (MST) protocol on the switches shown in the following figure.

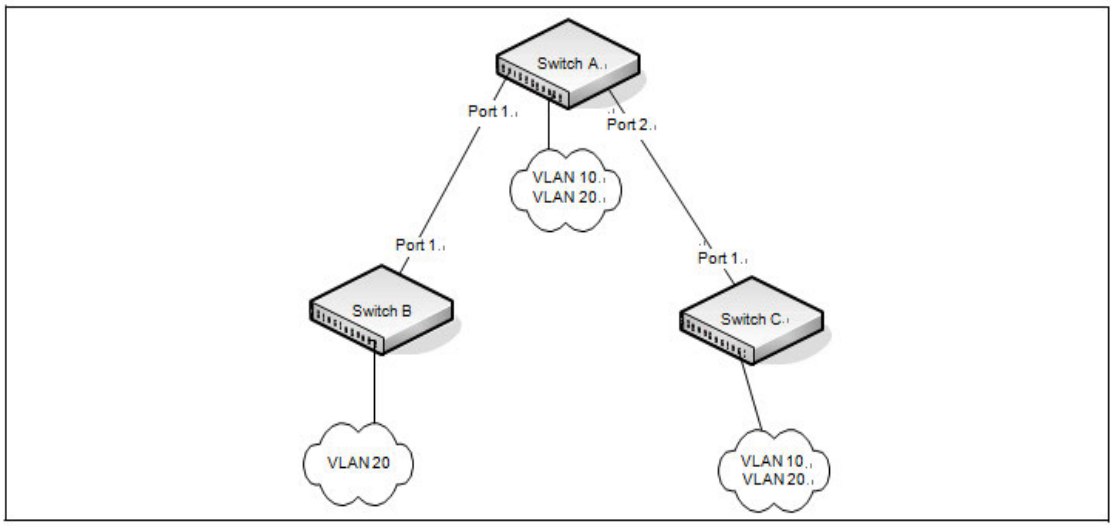


Figure 3-14: MSTP Configuration Example

To make multiple switches be part of the same MSTP region, make sure the STP operational mode for all switches is MSTP. Also, make sure the MST region name and revision level are the same for all switches in the region.

To configure the switches:

1. Create VLAN 10 and VLAN 20 (all switches).

Note: Even Switch B does not have any ports that are members of VLAN 10, this VLAN must be created to allow the formation of MST regions made up of all bridges that exchange the same MST Configuration Identifier. It is only within these MST Regions that multiple instances can exist.

```
(Switch) #config
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 10,20
(Switch) (Vlan)#exit
```

2. Set the STP operational mode to MSTP.

```
(Switch) #config
(Switch) (Config)#spanning-tree mode mstp
```

3. Create MST instance 10 and associate it to VLAN 10.

```
(Switch) (Config)#spanning-tree mst instance 10
(Switch) (Config)#spanning-tree mst vlan 10 10
```

4. Create MST instance 20 and associate it to VLAN 20.

```
(Switch) (Config)#spanning-tree mst instance 20
(Switch) (Config)#spanning-tree mst vlan 20 20
```

5. Change the region name so that all the bridges that want to be part of the same region can form the region.

```
(Switch) (Config)#spanning-tree configuration name NETGEAR
```

6. (Switch A only) Make Switch A the Regional Root for MSTI 1 by configuring a higher priority for MST ID 10.

```
(Switch) (Config)#spanning-tree mst priority 10 12288
```

7. (Switch A only) Change the priority of MST ID 20 to ensure Switch C is the Regional Root bridge for this MSTI.

```
(Switch) (Config)#spanning-tree mst priority 20 61440
```

8. (Switch C only) Change the priority of Switch C to force it to be the root bridge for MST 20.

```
(Switch) (Config)#spanning-tree mst priority 20 12288
(Switch) (Config)#exit
```

3.10.IGMP Snooping

IGMP Snooping is a layer-2 feature that allows the switch to dynamically add or remove ports from IP multicast groups by listening to IGMP join and leave requests. By “snooping” the IGMP packets transmitted between hosts and routers, the IGMP Snooping feature enables the switch to forward IP multicast traffic more intelligently and help conserve bandwidth.

Based on the IGMP query and report messages, the switch forwards traffic only to the ports that request the multicast traffic. This prevents the switch from broadcasting the traffic to all ports and possibly affecting network performance. The switch uses the information in the IGMP packets as they are being forwarded throughout the network to determine which segments should receive packets directed to the group address.

3.10.1. IGMP Snooping Querier

When PIM and IGMP are enabled in a network with IP multicast routing, the IP multicast router acts as the IGMP querier. However, if the IP-multicast traffic in a VLAN needs to be Layer 2 switched only, an IP-multicast router is not required. The IGMP Snooping Querier can perform the IGMP snooping functions on the VLAN.

Without an IP-multicast router on a VLAN, you must configure another switch as the IGMP querier so that it can send queries.

When the IGMP snooping querier is enabled, the IGMP snooping querier sends out periodic IGMP queries that trigger IGMP report messages from the switch that wants to receive IP multicast traffic. The IGMP snooping feature listens to these IGMP reports to establish appropriate forwarding.

3.10.2. Configuring IGMP Snooping

This example configures IGMP snooping on the switch to limit multicast traffic and to allow L2 multicast forwarding on a single VLAN. The IP-multicast traffic in VLAN 100 needs to be Layer 2 switched only, so the IGMP snooping querier is enabled on the switch to perform the IGMP snooping functions on the VLAN, if necessary. The switch can send queries even if it is not the IGMP snooping querier and will use 0.0.0.0 as the source IP address. This will not cause any disruption to the operation of external querier.

In this configuration, an IP-multicast router is not required.

In the following figure, the three hosts are connected to ports that are enabled for IGMP snooping and are members of VLAN 100. Port 24 is a trunk port and connects the switch to the data center, where the L3 multicast router is located.

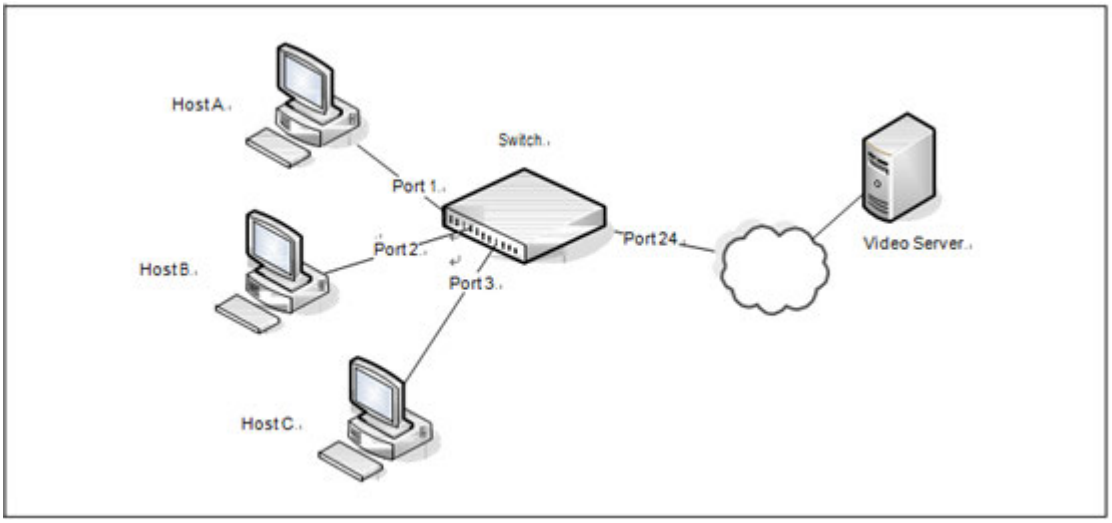


Figure 3-15: Switch with IGMP Snooping

To configure the switch:

1. Enable IGMP snooping globally.

```
(Switch) #configure
(Switch) (Config)#ip igmp snooping
```

2. Enable the IGMP snooping querier on the switch. If there are no other IGMP snooping queriers, this switch will become the IGMP snooping querier for the local network. If an external querier is discovered, this switch will not be a querier.

```
(Switch) (Config)#ip igmp snooping querier
```

3. Create VLAN 100

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 100
```

4. Enable IGMP snooping on VLAN 100.

```
(Switch) (Vlan)#set igmp 100
```

5. Enable the IGMP snooping querier on VLAN 100.

```
(Switch) (Config)#ip igmp snooping querier vlan 100
```

6. View the VLAN routing interface information.

```
(Switch) #show ip interface brief
                                                                                Netdir   Multi
Interface    State  IP Address      IP Mask                          Method   Bcast    CastFwd
----------   -----  --------------- --------------- --------------- -------- -------- --------
vlan 1       Down   0.0.0.0         0.0.0.0         Primary         None     Disable  Disable
vlan 100     Down   0.0.0.0         0.0.0.0         Primary         None     Disable  Disable
```

7. Configure an IP address for VLAN 100. This address will be used as the IGMP snooping querier address if this switch becomes the querier.

```
(Switch) #configure
(Switch) (Config)#interface vlan 100
(Switch) (if-vlan100)#ip address 192.168.10.2 255.255.255.0
(Switch) (if-vlan100)#exit
```

8. Specify the address to use as the source address for IGMP queries sent from any interface. The global querier address is the IP address of VLAN 100.

```
(Switch) (Config)#ip igmp snooping querier address 192.168.10.2
```

9. Enable IGMP snooping on ports 1–3.

```
(Switch) (Config)#interface range 0/1-0/3
(Switch) (Interface 0/1-0/3)#ip igmp snooping interfacemode
```

10. Configure ports 1–3 as members of VLAN 100.

```
(Switch) (Interface 0/1-0/3)#switchport allowed vlan add 100
(Switch) (Interface 0/1-0/3)#exit
```

11. Enable IGMP on port 24, and configure the port as a trunk port that connects to the data center switch.

```
(Switch) (Config)#interface 0/24
(Switch) (Interface 0/24)#ip igmp snooping interfacemode
(Switch) (Interface 0/24)#switchport allowed vlan add tagged 100
```

```
(Switch) (Interface 0/24)#exit
(Switch) (Config)#exit
```

12. Verify the IGMP snooping configuration.

```
(Switch) #show igmp snooping
Admin Mode..................................... Enable
Multicast Control Frame Count.................. 0
IGMP Snooping Router-Alert check.............. Disabled
Interfaces Enabled for IGMP Snooping.......... 0/1
                                               0/2
                                               0/3
                                               0/24
VLANs enabled for IGMP snooping............... 100
VLANs Block enabled for snooping.............. None

(Switch) #show igmp snooping querier vlan 100
VLAN 100 :IGMP Snooping querier status
---------------------------------------------
IGMP Snooping Querier VLAN Mode................ Enable
Querier Election Participate Mode.............. Disable
Querier VLAN Address........................... 0.0.0.0
Operational State.............................. Querier
Operational version............................ 2
Operational Max Resp Time...................... 10
```

After performing the configuration in this example, Host A sends a join message for multicast group 225.1.1.1. Host B sends a join message for group 225.1.1.2. Because IGMP snooping is enabled on the switch and on VLAN 100, the switch listens to the messages and dynamically adds Ports 1 and 2 to the multicast address table. Port 3 did not send a join message, so it does not appear in the table, as the following show command indicates.

```
(Switch) #show mac-address-table multicast

                                                                                Fwd
VLAN ID  MAC Address Source  Type  Description  InterfaceInterface

------- ---------------------- ------- -------------- ------------------

 100   01:00:5E:01:01:01    IGMP    Dynamic  Network Assist   0/1     0/1
 100   01:00:5E:01:01:02    IGMP    Dynamic  Network Assist   0/2     0/2
```

When the video server sends multicast data to group 225.1.1.1, Port 1 participates and receives multicast traffic, but Port 2 does not participate because it is a member of a different multicast group. Without IGMP snooping, all ports that are members of VLAN 100 would be flooded with traffic for all multicast groups, which would greatly increase the amount of traffic on the switch.

3.10.3. IGMPv3/SSM Snooping

IGMPv3 adds support for source filtering, which is the ability for a system to report interest in receiving packets **only** from specific source addresses, or from **all but** specific source addresses sent to a particular multicast address. This information is used by snooping switches to avoid delivering multicast packets from specific sources to networks where there are no interested receivers.

No additional configuration is required to enable IGMPv3/SSM snooping. It is enabled or disabled when snooping is enabled on a VLAN/interface. The forwarding database built using IGMPv3 reports is based on the Source IP address, the Multicast Group address, and VLAN. Consider the above configuration example. When Host A sends IGMPv3 IS_IN a report for Group 225.1.1.1 and Sources 192.168.10.1 and 192.168.20.1. As snooping is enabled globally on the switch and also on VLAN 100, two entries are added to MFDB so that multicast traffic with group IP = 225.1.1.1 and if Source Ip=192.168.10.1 or 192.168.20.1 is forwarded to port 1. All other multicast traffic destined to group 225.1.1.1 is dropped. The following command is used to display the SSM forwarding database.

```
(M4500-48XF8C) #show ip igmp snooping ssm entries
VLAN                                                          Source
ID   Group                  Source Ip              Filter Mode Interfaces
---- ---------------------- ---------------------- ----------- --------------
10   239.1.1.1              *                      include     0/7
10   239.1.1.1              80.1.1.1               include     0/5
10   239.1.1.1              80.1.1.1               exclude     0/7
```

3.11.SDVoE

SDVoE (Software Defined Video-over-Ethernet) is the latest high-performance, software-based AV-over-IP platform for control and distribution of audio and video over Ethernet & Fiber networks.

3.11.1. IGMP & IGMP Snooping Enhancements for IGMP V1 & V2

Note: All these enhancements are applicable for L2 multicast only, L3 multicast should behave as per the standard.

Note: This enhancement is applicable only to IGMP version 1 and version 2. Version 3 works very differently and is not part of this enhancement.

Note: This function only enabled on VLAN 1 by default and all interfaces in the system by default are part of VLAN 1.

When a VLAN is configured with SDVoE enabled, the following operation happens:

- All unknown data multicast will be blocked by default on VLAN 1.
- IGMP Snooping enabled by default for VLAN 1.
- IGMP snooping Fast-leave operation automatic:
  - Fast-leave is enabled on spanning-tree edge ports where encoders and decoders are connected.
  - Fast-leave is disabled on spanning-tree non-edge ports when connected to another switch.
- Flooding of IGMP (v1/v2) Join and Leave messages in a Switch.

  Note: This function is disable by default, user need to enable manually.

  - As per RFC 4541 all IGMP Join and Leave PDUs are processed by IGMP Snooping application and also forwarded through the mrouter port(s). Non-mrouter ports don't get to receive IGMP Join/Leave PDUs.
  - The intention of this part of the enhancement is to have a controlled way at a system level to flood an IGMP Join/Leave PDU received on a downstream port (from a host) to all other ports, given the boundary of the associated VLAN.
  - In a deployment where switches are connected in a star-like topology with one switch or stack connecting all others, this enables creating forwarding table entries across multiple switches and enables an Rx device on a switch to receive multicast data stream transmitted by a Tx device connected to another switch.
- Mrouter port to block all known (downstream hosts) and unknown Multicast data streams.
  - As per RFC 4541, a designated mrouter port (either detected dynamically or configured by user) forwards the following to the upstream router:
    1. All IGMP (v1/v2/v3) PDUs.
    2. All unknown Multicast streams – streams for which the switch has not received IGMP membership.
    3. All known Multicast streams – streams for which switch has received IGMP membership (from the hosts connected to it) and has its HW MFDB (Multicast Forwarding Data Base) table populated.
  - This enhancement aims at blocking the 2nd and 3rd points, only leaving 1st. Therefore, barring IGMP PDU forwarding, the mrouter port is to behave as any other host port. It forwards multicast data stream only if an IGMP membership (v1/v2) has been received through this port.
- Hardware forwarding of specific Multicast addresses.
  - The multicast addresses in the range 224.0.0.1 to 224.0.0.255 are considered as "Reserved" multicast addresses and are processed or forwarded by CPU, i.e., SW control plane. This enhancement is aimed to let a specific multicast addresses from this range and a few others to be forwarded by HW (ASICs) instead.

Here is the list of all Multicast destination addresses that are put under HW forwarding:

Address	Description
224.0.0.5	The Open Shortest Path First (OSPF). All OSPF Routers address is used to send Hello packets to all OSPF routers on a network segment.
224.0.0.6	The OSPF All Designated Routers ""(DR)"" address is used to send OSPF routing information to designated routers on a network segment.
224.0.0.9	The Routing Information Protocol (RIP) version 2 group address is used to send routing information to all RIP2-aware routers on a network segment.
224.0.0.18	Virtual Router Redundancy Protocol (VRRP).
224.0.0.107	Precision Time Protocol (PTP) version 2 peer delay measurement messaging.
224.0.0.251	Multicast DNS (mDNS) address.
224.0.0.252	Link-local Multicast Name Resolution (LLMNR) address.
224.0.1.1	Network Time Protocol clients listen on this address for protocol messages when operating in multicast mode.
224.0.1.129–132	Precision Time Protocol (PTP) version 1 messages (Sync, Announce, etc.) except peer delay measurement.
239.255.255.250	Simple Service Discovery Protocol address.
239.255.255.253	Service Location Protocol version 2 address.

Table 3-5: Specific Multicast addresses

When Switch receives multicast packet with any of the above address as destination address, switch installs a MFDB entry into HW so that all subsequent packets get forwarded at data plane. Switch places all physical ports and port-channels as part of the forwarding list of this entry for the VLAN packet has been received.

3.11.2. SDVoE Configuration Example

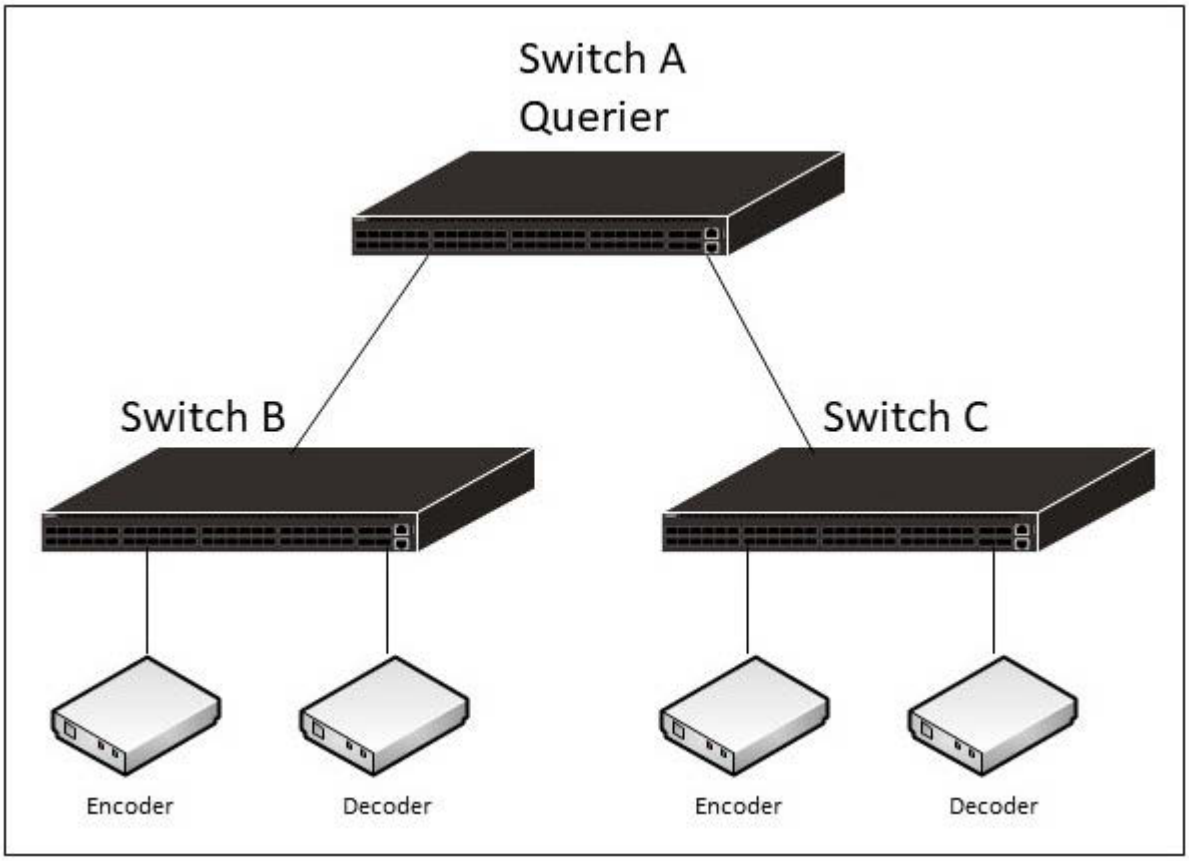


Figure 3-16: SDVoE Topology

The following example guide you how to setup SDVoE for VLAN 100.

1. Create VLAN 100 on Switch A, B and C, then enable SOVoE on both of switches.

```
(Switch) #configure

(Switch) (Config)#vlan database

(Switch) (Vlan)#vlan 100

(Switch) (Vlan)#exit

(Switch) (Config)#

(Switch) (Config)#sdvoe 100
```

2. Check the SDVoE status on Switch A, B and C.

```
(Switch) #show ip igmp snooping

Admin Mode..................................... Enable

Operation Mode................................. Enable
```

```
Multicast Control Frame Count.................. 0

IGMP Snooping Router-Alert check............... Disabled

Interfaces Enabled for IGMP Snooping........... None

VLANs enabled for IGMP snooping................ 1

                                                100

VLANs Block enabled for snooping............... None


(Switch) #show ip igmp snooping interface vlan 100

VLAN ID....................................... 100

IGMP Snooping Admin Mode...................... Enabled

Fast Leave Mode............................... Enabled

Flood IGMP Report and Leave PDU................ Disabled

Group Membership Interval (secs)............... 260

Max Response Time (secs)...................... 10

Multicast Router Block Mode.................... Enabled

Multicast Router Expiry Time (secs)............ 300

Report Suppression Mode....................... Disabled

Vlan Block Mode............................... Disabled
```

3. Configure IGMP Snooping Querier on Switch A.

```
(Switch-A) #configure

(Switch-A) (Config)#ip igmp snooping querier

(Switch-A) (Config)#ip igmp snooping querier vlan 100

(Switch-A) (Config)#ip igmp snooping querier vlan 100 address 192.168.10.253
```

4. Enable flooding IGMP PDUs on Switch A.

```
(Switch-A) #configure

(Switch-A) (Config)#vlan database

(Switch-A) (Vlan)#set igmp flood-report 100
```

5. Check the Querier Status.

```
(Switch-A) #show ip igmp snooping querier vlan 100

VLAN 100 :   IGMP Snooping querier status
```

```
-------------------------------------------

IGMP Snooping Querier VLAN Mode.............. Enable

Querier Election Participate Mode.............. Disable

Querier VLAN Address.......................... 192.168.10.253

Operational State............................. Querier

Operational version........................... 2

Operational Max Resp Time..................... 10
```

6. Check the flooding IGMP PDUs is enabled on Switch A.

```
(Switch) (Config)#show ip igmp snooping interface vlan 100

VLAN ID....................................... 100

IGMP Snooping Admin Mode...................... Enabled

Fast Leave Mode............................... Enabled

Flood IGMP Report and Leave PDU............... Enabled

Group Membership Interval (secs)............... 260

Max Response Time (secs)...................... 10

Multicast Router Block Mode................... Enabled

Multicast Router Expiry Time (secs)............ 300

Report Suppression Mode....................... Disabled

Vlan Block Mode............................... Disabled
```

3.12.MLD Snooping

3.12.1. MLD Snooping Configuration Example

When MLD Snooping is enabled on the switch, the switch can examine MLD packets and make forwarding decisions based on the MLD control packets content. You can configure the MLD snooping querier on the switch to support a subnet that does not have any multicast router interfaces. The MLD snooping querier periodically sends general MLD queries that the switch forwards through all ports in the VLAN.

There are three types of MLDv1 Message:

- Multicast Listener Query (Type = decimal 130)

  There are two subtypes of Multicast Listener Query messages:

  - General Query: used to learn which multicast addresses have listeners on an attached link.
  - Multicast-Address-Specific Query: used to learn if a particular multicast address has any listeners on an attached link.

- Multicast Listener Report (Type = decimal 131)
- Multicast Listener Done (Type = decimal 132)

There are three types of MLDv2 Queries Message:

- General Query: to learn which multicast address have multicast listeners.
- Multicast address specific query: to learn if a particular multicast address has any listeners.
- Multicast Address and Source Specific Queries: to learn if any of sources from the specified list for the particular multicast address has any listeners.

There are a number of different types of Multicast Address Records that can be included in an MLDv2 Report message:

- "Current State Record": is sent by a node in response to a Query received on an interface.
- "Filter Mode Change Record": is sent by a node whenever a local invocation of IPv6MulticastListen causes a change of the filter mode (i.e., a change from INCLUDE to EXCLUDE, or from EXCLUDE to INCLUDE) of the interface-level state entry for a particular multicast address, whether the source list changes at the same time or not.
- "Source List Change Record": is sent by a node whenever a local invocation of IPv6MulticastListen causes a change of source list that is *not* coincident with a change of filter mode, of the interface-level state entry for a particular multicast address.

3.12.1.1. MLD Snooping Configuration Example

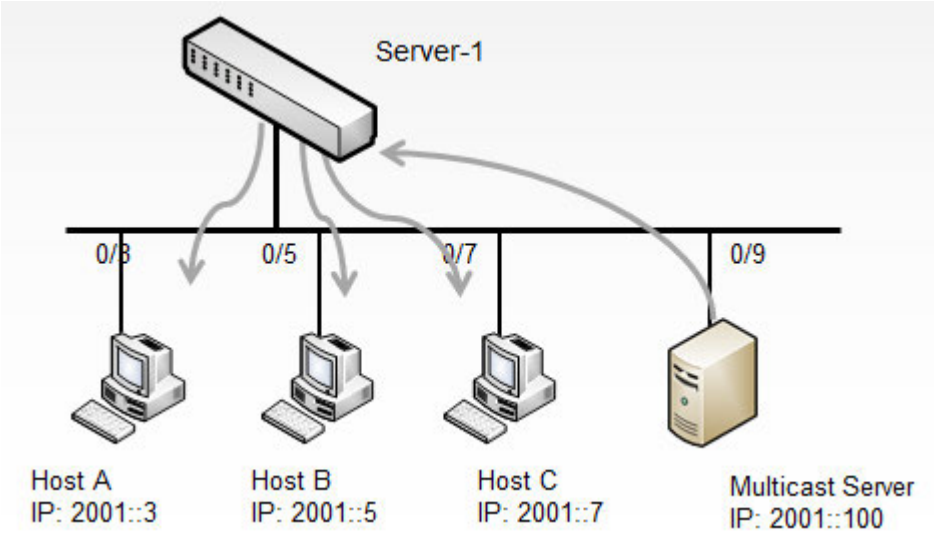


Figure 3-17: MLD Snooping Topology

Switch-1 MLD Snooping Configuration

Step 1. Enable MLD Snooping on admin mode

```
(Switch-1) (Config)#ipv6 mld snooping
```

Step 2. Enable MLD Snooping on VLAN 1.

```
(Switch-1) (Config)#vlan database
(Switch-1) (Vlan)#set mld 1
```

OR

Step 2. Enable MLD snooping on all interface and all VLANs

```
(Switch-1) (Config)#ipv6 mld snooping interfacemode all
```

OR

Step 2. Enable MLD snooping on specific interface 0/3-0/9

```
(Switch-1) (Config)#interface range 0/3-0/9
(Switch-1) (Interface 0/3-0/9)#ipv6 mld snooping interfacemode
```

3.12.1.2. MLD Snooping Verification Example

Switch-1 MLD Snooping Verification

Verify MLD snooping configuration on vlan 1

```
(Switch-1) (Config)#show ipv6 mld snooping interface vlan 1
VLAN ID..................................... 1
```

```
MLD Snooping Admin Mode..................... Enabled
Fast Leave Mode............................. Disabled
Group Membership Interval (secs)............ 260
Max Response Time (secs).................... 10
Multicast Router Expiry Time (secs)......... 0
Vlan Block Mode............................. Disabled
```

Verify MLD snooping configuration on interface 0/3

```
(Switch-1) (Config)#show ipv6 mld snooping interface 0/3
MLD Snooping Admin Mode..................... Enable
Fast Leave Mode............................. Disable
Group Membership Interval (secs)............ 260
Multicast Router Expiry Time (secs)......... 0
```

3.12.2. MLD Snooping First Leave Configuration Example

When MLD Fast Leave is enabled, a switch port will be removed immediately upon receiving an MLD done message as IGMP Leave message. The multicast clients leave from the multicast group quickly to reduce superfluous network traffic. According to the MLD v1 and v2 standard implementation, an MLD client may request to leave a multicast group by sending a Done message.

3.12.2.1. MLD Snooping Configuration

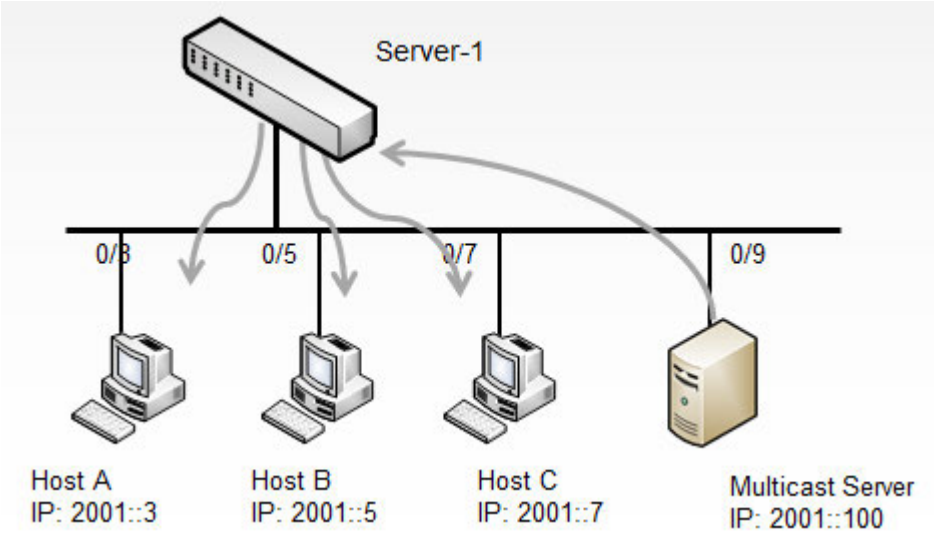


Figure 3-18: MLD Snooping Leave Configuration Topology

Step 1. Enable MLD Snooping on VLAN 1.

```
(Switch-1) (Config)#vlan database
(Switch-1) (Vlan)#set mld 1
(Switch-1) (Vlan)#set mld fast-leave 1
```

OR

Step 1. Enable MLD snooping on all interface and all VLANs

```
(Switch-1) (Config)#ipv6 mld snooping interfacemode all
(Switch-1) (Config)#ipv6 mld snooping fast-leave
```

OR

Step 1. Enable MLD snooping on specific interface 0/3-0/7

```
(Switch-1) (Config)#interface range 0/3-0/7
(Switch-1) (Interface 0/3-0/7)#ipv6 mld snooping interfacemode
(Switch-1) (Interface 0/3-0/7)#ipv6 mld snooping fast-leave
```

3.12.3. MLD Snooping Querier Configuration Example

If there is no mcast router in the network, then one of the switches should enable MLD snooping querier.

If there is no querier, then switch can't maintain mcast client information.

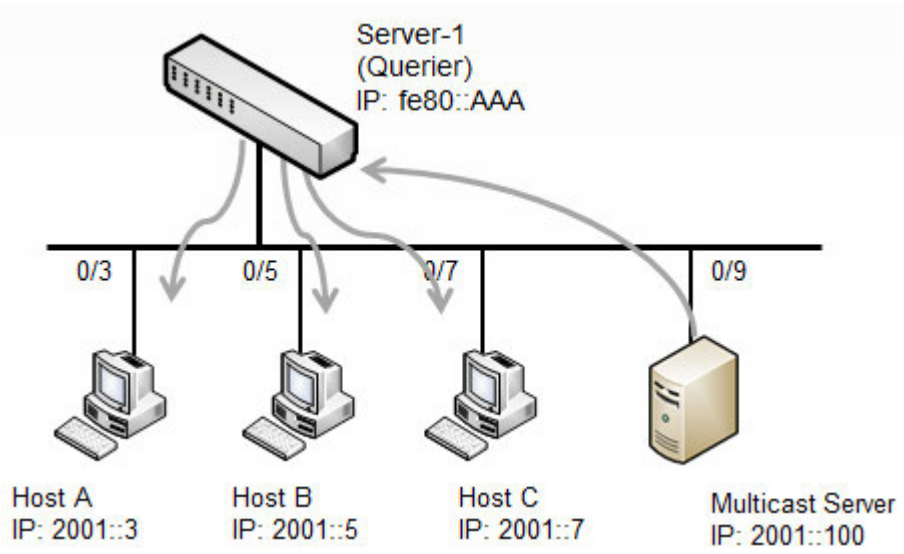


Figure 3-19: MLD Snooping Querier Configuration Example

MLD Snooping Querier Configuration

```
(Switch-1) (Config)#vlan database
(Switch-1) (Vlan)#set mld 1
(Switch-1) (Vlan)#exit
(Switch-1) (Config)#ipv6 mld snooping
```

```
(Switch-1) (Config)#ipv6 mld snooping querier address fe80::AAA
(Switch-1) (Config)#ipv6 mld snooping querier
(Switch-1) (Config)#ipv6 mld snooping querier vlan 1
```

Display MLD Snooping Querier detailed information.

```
(Switch-1) (Config)#show ipv6 mld snooping querier detail
VLAN ID Last Querier Address                     MLD Version
------- ---------------------------------------- -----------
Global MLD Snooping querier status
---------------------------------
MLD Snooping Querier Mode..................... Enable
Querier Address............................... fe80::aaa
MLD Version................................... 1
Querier Query Interval........................ 60
Querier Expiry Interval....................... 125
VLAN 1   :   MLD Snooping querier status
-------------------------------------------
MLD Snooping Querier VLAN Mode................ Enable
Querier Election Participate Mode............. Disable
Querier VLAN Address.......................... ::
Operational State............................. Querier
Operational version........................... 1
Operational Max Resp Time..................... 10
```

3.13.LLDP and LLDP-MED

LLDP is a standardized discovery protocol defined by IEEE 802.1AB. It allows stations residing on an 802 LAN to advertise major capabilities physical descriptions, and management information to physically adjacent devices allowing a network management system (NMS) to access and display this information.

LLDP is a one-way protocol; there are no request/response sequences. Information is advertised by stations implementing the transmit function, and is received and processed by stations implementing the receive function. The transmit-and-receive functions can be enabled/disabled separately on each switch port.

LLDP-MED is an extension of the LLDP standard. LLDP-MED uses LLDP's organizationally-specific Type-Length-Value (TLV) extensions and defines new TLVs that make it easier for a VoIP deployment in a wired or wireless LAN/MAN environment. It also makes mandatory a few optional TLVs from LLDP and recommends not transmitting some TLVs.

The TLVs only communicate information; these TLVs do not automatically translate into configuration. An external application may query the MED MIB and take management actions in configuring functionality.

LLDP and LLDP-MED are used primarily in conjunction with network management tools to provide information about network topology and configuration, and to help troubleshoot problems that occur on the network. The discovery protocols can also facilitate inventory management within a company.

LLDP and the LLDP-MED extension are vendor-neutral discovery protocols that can discover devices made by numerous vendors. LLDP-MED is intended to be used on ports that connect to VoIP phones. Additional applications for LLDP-MED include device location (including for Emergency Call Service/E911) and Power over Ethernet management.

3.13.1. LLDP and Data Center Application

DCBX uses TLV information elements over LLDP to exchange information, so LLDP must be enabled on the port to enable the information exchange.

3.13.2. Configuring LLDP

This example shows how to configure LLDP settings for the switch and to allow port 0/3 to transmit all LLDP information available.

To configure the switch:

1. Configure the transmission interval, hold multiplier, and reinitialization delay for LLDP PDUs sent from the switch.

```
(Switch) #configure
(Switch) (Config)#lldp timers interval 60 hold 5 reinit 3
```

2. Enable port 0/3 to transmit and receive LLDP PDUs.

```
(Switch) (Config)#interface 0/3
(Switch) (Interface 0/3)#lldp transmit (Switch) (Interface 0/3)#lldp receive
```

3. Enable port 0/3 to transmit management address information in the LLDP PDUs and to send topology change notifications if a device is added or removed from the port.

```
(Switch) (Interface 0/3)#lldp transmit-mgmt
(Switch) (Interface 0/3)#lldp notification
```

4. Specify the TLV information to be included in the LLDP PDUs transmitted from port 0/3.

```
(Switch) (Interface 0/3)#lldp transmit-tlv sys-name sys-desc sys-cap port-desc
```

5. Set the port description to be transmitted in LLDP PDUs.

```
(Switch) (Interface 0/3)#description “Test Lab Port”
```

6. Exit to Privileged EXEC mode.

```
(Switch) (Interface 0/3)# <CTRL + Z>
```

7. View global LLDP settings on the switch.

```
(Switch) #show lldp

LLDP Global Configuration

Transmit Interval..................... 60 seconds
Transmit Hold Multiplier.............. 5
Reinit Delay.......................... 3 seconds
Notification Interval................. 5 seconds
```

8. View summary information about the LLDP configuration on port 0/3.

```
(Switch) #show lldp interface 0/3
LLDP Interface Configuration

Interface  Link    Transmit  Receive   Notify    TLVs     Mgmt
---------  ------  --------  --------  --------  -------  ----
0/3        Down    Enabled   Enabled   Enabled   0,1,2,3    Y
TLV Codes: 0- Port Description,   1- System Name
           2- System Description, 3- System Capabilities
           4- Organization Specific
```

9. View detailed information about the LLDP configuration on port 0/3.

```
(Switch) #show lldp local-device detail 0/3

LLDP Local Device Detail

Interface: 0/3
Chassis ID Subtype: MAC Address
Chassis ID: 2C:60:0C:52:18:3F
Port ID Subtype: MAC Address
Port ID: 2C:60:0C:52:18:41
System Name: NETGEAR
System Description: LY8, Runtime Code 5.4.00.37, Linux 3.8.13-rt9, U-Boot 2010.12
(Oct 03 2014 - 14:38:07) – ONIE 2014.05.03-7
Port Description: Test Lab Port
System Capabilities Supported: bridge, router
System Capabilities Enabled: bridge
Management Address:
Type: IPv4
Address: 172.16.1.71
```

3.14.sFlow

sFlow is an industry standard technology for monitoring high-speed switched and routed networks. The software includes a built-in sFlow agent that can monitor network traffic on each port and generate sFlow data to an sFlow receiver (also known as a collector). sFlow helps to provide visibility into network activity, which enables effective management and control of network resources. The switch supports sFlow version 5.

As illustrated in the following figure, the sFlow monitoring system consists of sFlow Agents (such as a M4500 series switch) and a central sFlow receiver. sFlow Agents use sampling technology to capture traffic statistics from monitored devices. sFlow datagrams forward sampled traffic statistics to the sFlow Collector for analysis. You can specify up to eight different sFlow receivers to which the switch sends sFlow datagrams.

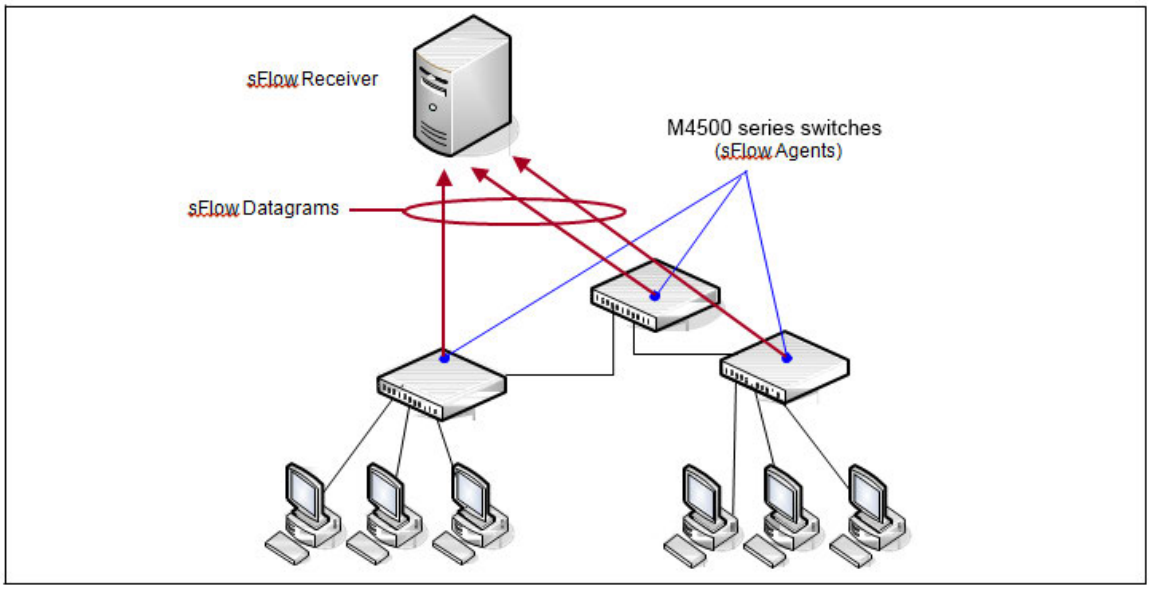


Figure 3-20: sFlow Architecture

The advantages of using sFlow are:

•It is possible to monitor all ports of the switch continuously, with no impact on the distributed switching performance.

•Minimal memory is required. Samples are not aggregated into a flow-table on the switch; they are forwarded immediately over the network to the sFlow receiver.

•The sFlow system is tolerant to packet loss in the network because statistical modeling means the loss is equivalent to a slight change in the sampling rate.

•sFlow receiver can receive data from multiple switches, providing a real-time synchronized view of the whole network.

•The receiver can analyze traffic patterns based on protocols found in the headers (e.g., TCP/IP, IPX, Ethernet, AppleTalk…). This alleviates the need for a layer 2 switch to decode and understand all protocols.

3.14.1. sFlow Sampling

The sFlow Agent uses two forms of sampling:

•Statistical packet-based sampling of switched or routed Packet Flows

•Time-based sampling of counters

Packet Flow Sampling and Counter Sampling are performed by sFlow Instances associated with individual Data Sources within an sFlow Agent. Both types of samples are combined in sFlow datagrams. Packet Flow Sampling creates a steady, but random, stream of sFlow datagrams that are sent to the sFlow Collector. Counter samples may be taken opportunistically to fill these datagrams.

To perform Packet Flow Sampling, an sFlow Sampler Instance is configured with a Sampling Rate. Packet Flow sampling results in the generation of Packet Flow Records. To perform Counter Sampling, an sFlow Poller Instance is configured with a Polling Interval. Counter Sampling results in the generation of Counter Records. sFlow Agents collect Counter Records and Packet Flow Records and send them as sFlow datagrams to sFlow Collectors.

3.14.1.1. Packet Flow Sampling

Packet Flow Sampling, carried out by each sFlow instance, ensures that any packet observed at a Data Source has an equal chance of being sampled, irrespective of the Packet Flow(s) to which it belongs.

Packet Flow Sampling is accomplished as follows:

•A packet arrives on an interface.

•The Network Device makes a filtering decision to determine whether the packet should be dropped.

•If the packet is not filtered (dropped), a decision is made on whether or not to sample the packet.

•A decision is made on whether or not to sample the packet.

The mechanism involves a counter that is decremented with each packet. When the counter reaches zero a sample is taken.

•When a sample is taken, the counter indicating how many packets to skip before taking the next sample is reset. The value of the counter is set to a random integer where the sequence of random integers used over time is the Sampling Rate.

3.14.1.2. Counter Sampling

The primary objective of Counter Sampling is to efficiently, periodically export counters associated with Data

Sources. A maximum Sampling Interval is assigned to each sFlow instance associated with a Data Source.

Counter Sampling is accomplished as follows:

•sFlow Agents keep a list of counter sources being sampled.

•When a Packet Flow Sample is generated the sFlow Agent examines the list and adds counters to the sample datagram, least recently sampled first. Counters are only added to the datagram if the sources are within a short period, 5 seconds say, of failing to meet the required Sampling Interval.

•Periodically, say every second, the sFlow Agent examines the list of counter sources and sends any counters that must be sent to meet the sampling interval requirement.

The set of counters is a fixed set.

3.14.2. Configuring sFlow

This example shows how to configure the switch so that ports 10-15 and port 23 send sFlow datagrams to an sFlow receiver at the IP address 192.168.20.34. The receiver owner is receiver1, and the timeout is 100000 seconds. A counter sample is generated on the ports every 60 seconds (polling interval), and 1 out of every

8192 packets is sampled. To configure the switch:

1. Configure information about the sFlow receiver.

```
(Switch) #configure
(Switch) (Config)#sflow receiver 1 ip 192.168.20.34
(Switch) (Config)#sflow receiver 1 owner receiver1 timeout 100000
```

2. Configure the polling and sampling information for ports 10–15.

```
(Switch) (Config)#interface range 0/10-0/15
(Switch) (Interface 0/10-0/15)#sflow poller 1
(Switch) (Interface 0/10-0/15)#sflow poller interval 60
(Switch) (Interface 0/10-0/15)#sflow sampler 1
(Switch) (Interface 0/10-0/15)#sflow sampler rate 8192
(Switch) (Interface 0/10-0/15)#exit
```

3. Configure the polling and sampling information for port 23.

```
(Switch) (Config)#interface 0/23
(Switch) (Interface 0/23)#sflow poller 1
(Switch) (Interface 0/23)#sflow poller interval 60
(Switch) (Interface 0/23)#sflow sampler 1
(Switch) (Interface 0/23)#sflow sampler rate 8192
(Switch) (Interface 0/23)#exit
```

4. Verify the configured information.

```
(Switch) #show sflow receivers 1
Receiver Index.............................. 1
Owner String................................ receiver1
Time out.................................... 99400
```

```
IP Address:..................................... 192.168.20.34

Address Type.................................... 1

Port............................................ 6343

Datagram Version................................ 5

Maximum Datagram Size.......................... 1400 (Switch) #show sflow pollers

PollerReceiver  Poller

Data SourceIndexInterval

-------------------------

0/10160

0/11160

0/12160

0/13160

0/14160

0/15 1   60

0/23 1   60


(Switch) #show sflow samplers

  Sampler      Receiver  Packet          Max Header
  Data Source  Index     Sampling Rate   Size
  -----------  -------   -------------   ----------
  0/10         1         8192            128
  0/11         1         8192            128
  0/12         1         8192            128
  0/13         1         8192            128
  0/14         1         8192            128
  0/15         1         8192            128
  0/23         1         8192            128
```

3.15.Link Dependency

The following commands configure a link-dependency group.

1. Create a link dependency group with group ID 1. This command also configures whether the downstream interfaces should mirror or invert the status of upstream interfaces. The **action up** command causes the downstream interfaces to be up when no upstream interfaces are down.

```
(Switch) #configure

(Switch) (Config)#link state group 1 action down
```

2. Configure ports as link-dependency group members. Port 0/8 is configured as an upstream member of the group and ports 0/3 and 0/5 are configured as downstream members. The state of downstream members is dependent on the state of the upstream member.

Circular dependencies are not allowed. An interface that is defined as an upstream interface cannot also be defined as a downstream interface in the same link state group. An interface that is defined as an upstream interface cannot also be defined as a downstream interface in a different link state group, when such configuration creates a circular dependency between groups.

Caution! Adding an interface as a downstream port brings the interface down until an upstream interface is added to the group. The link status will then follow the interface specified in the upstream command. To avoid bringing down interfaces, configure the upstream port prior to configuring the downstream ports.

```
(Switch) (Config)#interface 0/8
(Switch) (Interface 0/8)#link state group 1 upstream
(Switch) (Interface 0/8)#exit

(Switch) (Config)#interface range 0/3, 0/5
(Switch) (Interface 0/3,0/5)#link state group 1 downstream
(Switch) (Interface 0/3,0/5)#exit
```

To view link dependency settings for all groups or for the specified group, along with the group state, use the commands **show link state group [***group_id***]** and **show link state group** *group-id* **detail**.

3.16.FIP Snooping

FIP snooping is a frame inspection method used by the FIP Snooping Bridge to monitor FIP frames and apply policies based on the L2 header information in those frames, following recommendations in Annex C of FC_BB_5 Rev 2.00.

FIP Snooping enables the following features:

- Auto-configuration of Ethernet ACLs based on information in the Ethernet headers of FIP frames.
- Emulation of fiber channel (FC) point-to-point links within the DCB Ethernet network.
- Enhanced FCoE security/robustness by preventing FCoE MAC spoofing.

The FIP Snooping Bridge solution is intended for use only at the edge or perimeter of the switched network and not on an interior switch.

To configure FIP snooping:

1. For ports connected to CNAs/ENodes, enable LLDP and DCBX and configure them as DCBX *auto-down* ports. In this example, the ports connected to the CNAs/ENodes are ports 0/9 and 0/10.

```
(Switch) #config
(Switch) (Config)#interface range 0/9-0/10
(Switch) (Interface 0/9-0/10)#lldp transmit
(Switch) (Interface 0/9-0/10)#lldp receive
(Switch) (Interface 0/9-0/10)#lldp dcbx port-role auto-down
(Switch) (Interface 0/9-0/10)#exit
```

2. For ports connected to the FCoE Forwarders (FCFs), enable LLDP and DCBX and configure these ports as DCBX *auto-up* ports. In this example, the port connected to the FCF is port 0/11.

```
(Switch) (Config)#interface 0/11
(Switch) (Interface 0/11)#lldp transmit
(Switch) (Interface 0/11)#lldp receive
(Switch) (Interface 0/11)#lldp dcbx port-role auto-up
(Switch) (Interface 0/11)#exit
```

3. In Global Config mode, configure one-to-one global dot1p mapping.

```
(Switch) (Config)#queue cos-map all 0 0
(Switch) (Config)#queue cos-map all 1 1
(Switch) (Config)#queue cos-map all 2 2
(Switch) (Config)#queue cos-map all 3 3
(Switch) (Config)#queue cos-map all 4 4
(Switch) (Config)#queue cos-map all 5 5
(Switch) (Config)#queue cos-map all 6 6
(Switch) (Config)#queue cos-map all 7 7
(Switch) (Config)#exit
```

4. Create the FCoE VLAN. In this example, FCoE VLAN ID is 1000.

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 1000
(Switch) (Vlan)#exit
```

5. Add VLAN 1000 membership to the ports connected to CNAs and FCF. Enable VLAN tagging on these ports for FCoE VLAN using below interface commands.

```
(Switch) #config
(Switch) (Config)#interface range 0/9-0/11
(Switch) (Interface 0/9-0/11)#switchport allowed vlan add tagged 1000
(Switch) (Interface 0/9-0/11)#exit
(Switch) (Config)#exit
```

6. Enable FIP snooping in FCoE VLAN 1000.

```
(Switch) #configure
(Switch) (Config)#feature fip-snooping
(Switch) (Config)#vlan 1000
(Switch) (Config)(Vlan 1000)#fip-snooping enable
(Switch) (Config)(Vlan 1000)#exit
(Switch) (Config)#exit
```

7. Configure FCF facing ports using below interface command. By default, FIP snooping ports are configured as host/ENode mode.

```
(Switch) #configure
(Switch) (Config)#interface 0/11
(Switch) (Interface 0/11)#fip-snooping port-mode fcf
(Switch) (Interface 0/11)#exit
(Switch) (Config)#exit
```

The following command sample shows the configuration script for the FIP snooping switch configured in the example. Two interfaces (0/9 and 0/10) are connected to CNAs, and 0/11 is connected to the FCF.

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 1000
(Switch) (Vlan)#exit

(Switch) #configure
(Switch) (Config)#feature fip-snooping
(Switch) (Config)#vlan 1000
(Switch) (Config)(Vlan 1000)#fip-snooping enable
(Switch) (Config)(Vlan 1000)#exit

(Switch) (Config)#queue cos-map all 0 0
(Switch) (Config)#queue cos-map all 1 1
(Switch) (Config)#queue cos-map all 2 2
(Switch) (Config)#queue cos-map all 3 3
```

```
(Switch) (Config)#queue cos-map all 4 4
(Switch) (Config)#queue cos-map all 5 5
(Switch) (Config)#queue cos-map all 6 6
(Switch) (Config)#queue cos-map all 7 7

(Switch) (Config)#interface 0/9
(Switch) (Interface 0/9)#description 'NETGEAR CNA'
(Switch) (Interface 0/9)#switchport allowed vlan add tagged 1000
(Switch) (Interface 0/9)#switchport priority 3
(Switch) (Interface 0/9)#lldp transmit
(Switch) (Interface 0/9)#lldp receive
(Switch) (Interface 0/9)#lldp dcbx port-role auto-down
(Switch) (Interface 0/9)#exit

(Switch) (Config)#interface 0/10
(Switch) (Interface 0/10)#description 'NETGEAR1 CNA'
(Switch) (Interface 0/10)#switchport allowed vlan add tagged 1000
(Switch) (Interface 0/10)#switchport priority 3
(Switch) (Interface 0/10)#lldp transmit
(Switch) (Interface 0/10)#lldp receive
(Switch) (Interface 0/10)#lldp dcbx port-role auto-down
(Switch) (Interface 0/10)#exit

(Switch) (Config)#interface 0/11
(Switch) (Interface 0/11)#description 'NETGEAR2-FCF Facing'
(Switch) (Interface 0/11)#switchport allowed vlan add tagged 1000
(Switch) (Interface 0/11)#switchport priority 3
(Switch) (Interface 0/11)#fip-snooping port-mode fcf
(Switch) (Interface 0/11)#lldp transmit
(Switch) (Interface 0/11)#lldp receive
(Switch) (Interface 0/11)#lldp dcbx port-role auto-up
(Switch) (Interface 0/11)#exit
(Switch) (Config)#exit
```

3.17.ECN

Explicit Congestion Notification (ECN) is defined in RFC 3168. Conventional TCP networks signal congestion by dropping packets. A Random Early Discard scheme provides earlier notification than a tail drop scheme by dropping packets already queued for transmission. ECN marks congested packets that would otherwise have been dropped and expects an ECN-capable receiver to signal congestion back to the transmitter without the need to retransmit the packet that would have been dropped. For TCP, this means that the TCP receiver signals a reduced window size to the transmitter but does not request retransmission of the CE marked packet.

ECN uses the two least significant bits of Diffserv field (TOS octet in IPv4/Traffic Class octet in IPv6) and codes them as follows:

00: Non ECN-Capable Transport – Non-ECT

10: ECN Capable Transport – ECT(0)

01: ECN Capable Transport – ECT(1)

11: Congestion Encountered – CE

ECN-capable hosts communicate support for ECN via two flags in the TCP header:

•ECN-Echo (ECE)

•Congestion Window Reduced (CWR)

WRED considers packets for early discard only when the number of packets queued for transmission on a port exceeds the relevant minimum WRED threshold. The green, yellow, red thresholds operate on TCP packets. The fourth threshold operates on non-TCP packets.

When ECN is enabled and congestion is experienced, TCP packets that are marked ECN Capable that are queued for transmission and are selected for discarded by WRED, are instead marked CE and transmitted. This includes packets that exceed the WRED upper threshold. If the switch experiences severe congestion (no buffers available), then packets are discarded.

WRED considers packets for early discard only when the number of packets queued for transmission on a port exceeds the relevant minimum WRED threshold. Four thresholds are available for configuration. The green, yellow, and red thresholds operate on TCP packets. The fourth threshold operates on non-TCP packets.

When ECN is enabled and congestion is experienced, packets that are marked ECN-capable, are queued for transmission, and are randomly selected for discard by WRED are instead marked CE and are transmitted rather than dropped. This includes packets that exceed the WRED upper threshold. If the switch experiences severe congestion (no buffers available), then packets are discarded.

The switch supports ECN capability as part of the WRED configuration process. Eligible packets are marked by hardware based on the WRED configuration. The network operator can configure any CoS queue to operate in ECN marking mode and can configure different discard thresholds for each color.

3.17.1. Enabling ECN in Microsoft Windows

On many current Windows implementations, ECN capability is enabled via the **netsh** command as follows:

```
netsh int tcp set global ecncapability=enabled


The capability can be verified with the following command:


netsh int tcp show global.


An example is shown below:


C:\Users\user1>Netsh int tcp set global ecncapability=enabled

Ok.

C:\Users\user1>netsh int tcp show global

Querying active state...


TCP Global Parameters

----------------------------------------------

Receive-Side Scaling State : enabled

Chimney Offload State : automatic

NetDMA State : enabled

Direct Cache Acess (DCA) : disabled Receive Window Auto-Tuning Level : normal Add-On Congestion
Control Provider : none ECN Capability : enabled

RFC 1323 Timestamps : disabled
```

In Windows Server 2012, DCTCP is self-activating based on the RTT of TCP packets. No user management is required. Use the PowerShell cmdlet **Get-NetTcpConnection** to verify DCTCP operation.

3.17.2. Example 1: SLA Example

The following example configures simple meter and a trTCM meter in support of a network SLA. The SLA classes are segregated by CoS class.

1. Define a class-map so that all traffic will be in the set of traffic “cos-any”.

```
(Switch) (Config)#class-map match-all cos-any ipv4

(Switch) (Config-classmap)#match any

(Switch) (Config-classmap)#exit
```

2. Define a class-map such that all traffic with a Cos value of 1 will be in the set of traffic “cos1”. This will be used as a conform-color class map. Conform-color class maps must be one of CoS, secondary CoS, DSCP, or IP precedence.

```
(Switch) (Config)#class-map match-all cos1 ipv4
(Switch) (Config-classmap)#match cos 1
(Switch) (Config-classmap)#exit
```

3. Define a class-map such that all IPv4 traffic with a CoS value of 0 will be in the set of traffic “cos0”. This will be used as a conform-color class map. Conform-color class maps must be one of CoS, secondary CoS, DSCP, or IP precedence.

```
(Switch) (Config)#class-map match-all cos0 ipv4
(Switch) (Config-classmap)#match cos 0
(Switch) (Config-classmap)#exit
```

4. Define a class-map such that all TCP will be in the set of traffic “TCP”. This will be used as a base color class for metering traffic.

```
(Switch) (Config)#class-map match-all tcp ipv4
(Switch) (Config-classmap)#match protocol tcp
(Switch) (Config-classmap)#exit
```

5. Define a policy-map to include packets matching class “cos-any” (IPv4). Ingress IPv4 traffic arriving at a port participating in this policy will be assigned red or green coloring based on the metering.

```
(Switch) (Config)#policy-map simple-policy in
(Switch) (Config-policy-map)#class cos-any
```

6. Create a simple policer in color blind mode. Packets below the committed information rate (CIR) or committed burst size (CBS) are assigned drop precedence “green”. Packets that exceed the CIR (in Kbps) or CBS (in Kbytes) are colored “red”. Both the conform and violate actions are set to transmit as WRED is used to drop packets when congested.

```
(Switch) (Config-policy-classmap)#police-simple 10000000 64 conform-action transmit violate-
action transmit
(Switch) (Config-policy-classmap)#exit
(Switch) (Config-policy-map)#exit
```

7. Define a policy-map in color aware mode matching class “cos-any” (IPv4). Ingress IPv4 traffic arriving at a port participating in this policy will be assigned green, yellow, or red coloring based on the meter.

```
(Switch) (Config)#policy-map two-rate-policy in
(Switch) (Config-policy-map)#class tcp
```

8. Create a two-rate policer per RFC 2698. The CIR value is 800 Kbps and the CBS is set to 96 Kbytes. The PIR is set to 950 Kbps and the PBS is set to 128 Kbytes. Color-aware processing is enabled via the **conform-color** command (i.e., any packets not in cos 0 or 1 are pre-colored “red.” Packets in cos 0 are pre- colored yellow. Packets in cos 1 are pre-colored green. Pre-coloring gives greater bandwidth to cos 1 packets, as they are initially subject to the CIR/CBS limits. Packets in CoS 0 are subject to the PIR limits. Based on the CIR/CBD, the PIR/PBS, and the conform, exceed, and violate actions specified below.

TCP packets with rates less than or equal to the CIR/CBS in class cos 1 are conforming to the rate (green). These packets will be dropped randomly at an increasing rate between 0–3% when the outgoing interface is congested between 80 and 100%.

TCP packets with rates above the CIR/CBS and less than or equal to PIR/PBS in either class cos 1 or class cos 2 are policed as exceeding the CIR (yellow). These packets will be dropped randomly at an increasing rate between 0–5% when the outgoing interface is congested between 70 and 100%. TCP packets with rates higher than the PIR/PBS or which belong to neither class cos 1 nor class cos 2 are violating the rate (red).

These packets will be dropped randomly at an increasing rate between 0–10% when the outgoing interface is congested between 50 and 100%.

Non-TCP packets in CoS queue 0 or 1 will be dropped randomly at an increasing rate between 0–15% when the outgoing interface is congested between 50 and 100%.

```
(Switch) (Config-policy-classmap)#police-two-rate 800 96 950 128 conform-action transmit exceed-
action transmit violate-action transmit
(Switch) (Config-policy-classmap)#conform-color cos1 exceed-color cos0
(Switch) (Config-policy-classmap)#exit
(Switch) (Config-policy-map)#exit
```

9. Enable WRED drop on traffic classes 0 and 1.

```
(Switch) (Config)#queue cos-queue random-detect 0 1
```

10. Set the exponential-weighting-constant. The exponential weighting constant smooths the result of the average queue depth calculation by the function:

average depth = (previous queue depth * (1-1/2^n)) + (current queue depth * 1/2^n).

The average depth is used in calculating the amount of congestion on a queue. Because the instantaneous queue depth fluctuates rapidly, larger values of the weighting constant cause the average queue depth value to respond to changes more slowly than smaller values.

```
(Switch) (Config)#random-detect exponential-weighting-constant 4
```

11. Configure the queue parameters for traffic class 0 and 1. We set the minimum threshold and maximum thresholds to 80–100% for green traffic, 70–100% for yellow traffic, and 50–100% for red traffic. Non-TCP traffic drops in the 50–100% congestion range. Green traffic is dropped at a very low rate to slowly close the TCP window. Yellow and red traffic are dropped more aggressively.

```
(Switch) (Config)#random-detect queue-parms 0 1 min-thresh 80 70 50 50 max-thresh 100 100 100
100 drop-prob 3 5 10 15
```

12. Assign the color policies to ports. The metering policies are applied on ingress ports.

```
(Switch) (Config)#interface 0/22
(Switch) (Interface 0/22)#service-policy in simple-policy
(Switch) (Interface 0/22)#exit
(Switch) (Config)#interface 0/23
(Switch) (Interface 0/23)#service-policy in two-rate-policy
(Switch) (Interface 0/23)#exit
```

3.17.3. Example 2: Data Center TCP (DCTCP) Configuration

This example globally configures a switch to utilize ECN marking of green packets queued for egress on CoS queues 0 and 1, using the DCTCP threshold as it appears in “DCTCP: Efficient Packet Transport for the Commoditized Data Center” (Alizadeh, Greenberg, Maltz, Padhye, Patel, Prabhakar, Sengupta, and Sridharan, 2010.)

In the first line of the following configuration, the first integer after the **minthresh** keyword configures green-colored Congestion Enabled TCP packets in CoS queues 0 and 1 that exceed the WRED threshold (13% or

~38 Kbytes) to mark packets as Congestion Experienced. The first integer after the **max-thresh** parameter configures the upper threshold for green-colored TCP packets to the same value as the **min-thresh** threshold. This causes the switch to mark all ECN-capable queued packets as Congestion Experienced when the threshold is reached or exceeded. TCP packets without ECN capability bits set are dropped according to the normal WRED processing when the threshold is exceeded. Packets on other CoS queues are handled in the standard manner (i.e., are tail-dropped) when insufficient buffer is available.

Yellow and red packet configuration (second and third threshold parameters) are kept at the defaults, as no metering to reclassify packets from green to yellow or red is present. The last threshold parameter configures non-TCP packets in CoS queues 0 and 1 to be processed with the WRED defaults. The **ecn** keyword enables ECN marking of ECN-capable packets on CoS queues 0 and 1. The weighting constant is set to 0 in the second line of the configuration, as described in the DCTCP paper cited above. Finally, CoS queues 0 and 1 are configured for WRED in the last line of the configuration.

```
(Switch) #config

(Switch) (Config)#random-detect queue-parms 0 1 min-thresh 13 30 20 100 max-thresh 13 90 80 100
drop-prob 100 10 10 10 ecn

(Switch) (Config)#random-detect exponential-weighting-constant 0

(Switch) (Config)#queue cos-queue random-detect 0 1
```

3.18.Storm Control

When Layer 2 frames are forwarded, broadcast, unknown unicast, and multicast frames are flooded to all ports on the relevant virtual local area network (VLAN). The flooding occupies bandwidth, and loads all nodes connected on all ports. Storm control limits the amount of broadcast, unknown unicast, and multicast frames accepted and forwarded by the switch.

Per-port and per-storm control type (broadcast, multicast, or unicast), the storm control feature can be configured to automatically shut down a port when a storm condition is detected on the port; or to send a trap to the system log. When configured to shut down, the port is put into a diagnostic-disabled state. The user must manually re-enable the interface for it to be operational. When configured to send a trap, the trap is sent once in every 30 seconds. When neither action is configured, the switch rate-limits the traffic when storm conditions occur.

3.18.1. Storm Control Configuration Example

Assume PC-A transmits wire speed broadcast traffic continuously, enable storm-control features on port connect PC-A can decline heavy traffic into whole network.

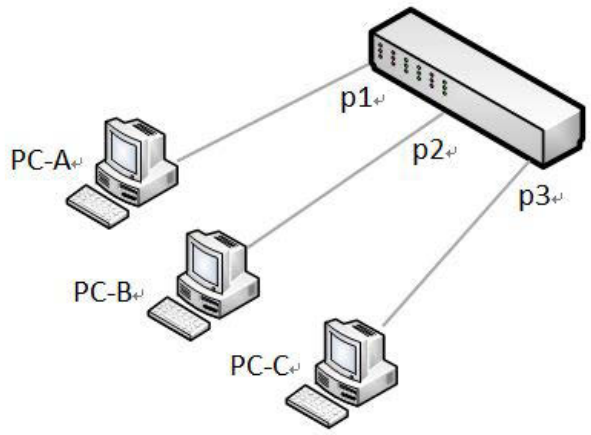


Figure 3-21: Storm Control Topology

Enabling broadcast storm control (Multicast / Unicast is the same configure as broadcast)

1. Enter interface configuration mode and enable storm-control.

```
(Switch) #configure
(Switch) (config) #interface 0/1
(Switch) (Interface 0/1) #storm-control broadcast
```

2. Configure broadcast storm control as 20 percent of port speed for broadcast traffic.

```
(Switch) (Interface 0/1) #storm-control broadcast level 20
```

3. Configure the action to be taken when a storm is detected. The shutdown keyword sets the port to error-disable state.

```
(Switch) (Interface 0/1) #storm-control broadcast action shutdown
```

4. Verify the configuration.

```
(Switch) #show storm-control all

       Bcast   Bcast    Bcast      Mcast   Mcast    Mcast      Ucast   Ucast    Ucast
 Intf  Mode    Level    Action     Mode    Level    Action     Mode    Level    Action
------ ------- -------- ---------- ------- -------- ---------- ------- -------- ---------- ---
0/1    Enable  20%      Shutdown   Disable 5%       None       Disable 5%       None
0/2    Disable 5%       None       Disable 5%       None       Disable 5%       None
0/3    Disable 5%       None       Disable 5%       None       Disable 5%       None
```

3.19.Jumbo Frames

Jumbo frames enable transporting data in fewer frames to ensure less overhead, lower processing time, and fewer interrupts. The maximum transmission unit (MTU) size is configurable per-port.

3.19.1. Jumbo Frame Configuration Example

1. Changes the MTU size for interface 0/1 on the switch.

```
(Switch) #configure
(Switch) (Config)#interface 0/1
(Switch) (Interface 0/1)#mtu 9000
(Switch) (Interface 0/1)#exit
```

2. Verify the configuration.

```
(Switch) #show interface counters detailed 0/1

......

Max Frame Size.............................. 9000
```

3.20.Port-Backup

Port backup are a pair of a Layer 2 interfaces where one interface is configured to active as a backup to the other. The feature provides an alternative solution for redundancy.

In the following figure, Switch A is connected to Switch B and Switch C via p1 and p2. Because these interfaces are configured as port-backups, only one of the interfaces is forwarding traffic; the other is in standby mode. If port 1 is the active link, it begins forwarding traffic between port 1 and Switch B; the link between port 2 (the backup link) and Switch C is not forwarding traffic. If port 1 goes down, port 2 comes up and starts forwarding traffic to Switch C.

When port 1 comes back up, it goes into standby mode and does not forward traffic before failback-time expired; port 2 continues forwarding traffic.

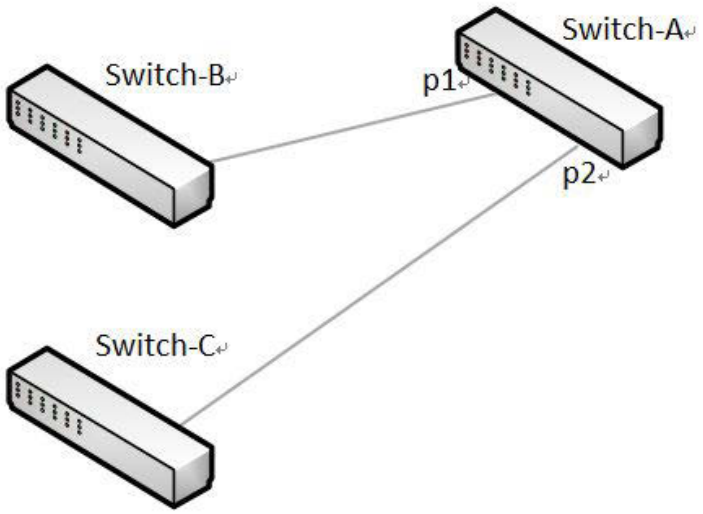


Figure 3-22: Port-Backup Topology

3.20.1. Port-Backup Configuration Example

1. Create port-backup group before enable this feature.

```
(Switch) #configure
(Switch) (Config)#port-backup group
Port backup group 1 is created
```

2. Assign P1 to port-backup group as active port and P2 as backup port.

```
(Switch) (Config)#interface 0/1
(Switch) (Interface 0/1)#port-backup group 1 active
(Switch) (Interface 0/1)#interface 0/2
(Switch) (Interface 0/2)#port-backup group 1 backup
(Switch) (Interface 0/2)#exit
```

3. Enable port-backup group 1.

```
(Switch) (Config)#port-backup group enable 1
```

4. Enable port-backup admin mode.

```
(Switch) (Config)#port-backup
```

When current active port change from “active port” to “backup” port, the mac-move-update function will let backup port to send out the source addresses to peer link to make sure the learned MAC addresses on previous active port peer link can learned on backup port’s peer link.

5. Enable mac address table move update feature.

```
(Switch) (Config)#port-backup group 1 mac-move-update
```

6. Verify the configuration.

```
(Switch) #show port-backup


Admin Mode: Enable
Group Mode MAC Update Failback Active Port  Backup Port  Current Active Port
----- ---- ---------- -------- ------------ ------------ --------------------
1      En. Enable     60(sec)  0/1           0/2
```

3.21.PTP End-to-End Transparent Clock

Precision Time Protocol (PTP, IEEE 1588) is a protocol that enables precise synchronization of clocks with a sub-microsecond accuracy across a packet-based network. PTP lets network devices of different precision and resolution synchronize to a grandmaster clock through an exchange of packets across the network. The

switch supports PTP end-to-end transparent clock, which is enabled by default, both globally and at the port level. Note that the switch itself is not affected by PTP.

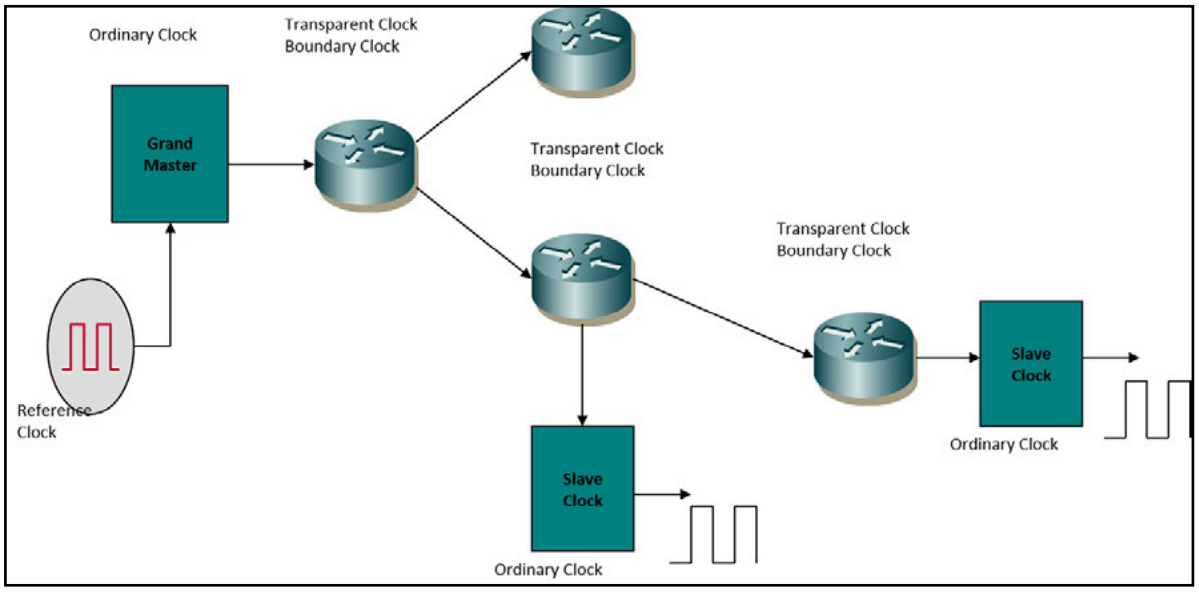


Figure 3-23: Port-Backup Topology

3.21.1. PTP Time Stamp Operation

The IEEE 1588 protocol (PTP) requires protocol PTP packets to be time-stamped when they pass through the network.

Two types of time stamp operation exist:

- Time stamp recording. The switch records the time that a PTP event packet is received or transmitted. The record is available to the local CPU.

- Time stamp correction. The PTP packets are modified as they pass through a network device. The PTP packets include a correction field that can carry clock corrections for residence times of a transparent clock, path delay for peer-to-peer transparent clocks, and so on. The timestamp value is deducted from the correction field when the PTP packet enters the device and is added to the correction field when the packet leaves the device. The result is that the correction field is updated with the period that the packet spends in the device.

Two types of hardware time stamp operation exist, one of which is supported by the M4500 switches:

- **1-step hardware time stamp**. The hardware supports corrections, allowing the PTP packets to be modified as they pass through the device. The M4500 series switches support 1-step hardware time stamps only:
  The switch supports PTPv2 packets with a destination MAC address that is set to 01:1B:19:00:00:00. The switch does not support PTPv1 packets and drops those packets. The time stamp counter is 32 bits in length.

- **2-step hardware time stamp**. The hardware supports recording of time stamps of the PTP event packets. This time stamp is then retrieved by the local CPU on a device that supports the PTP firmware and is sent in a separate message. The M4500 series switches do not support 2-step hardware time stamps.

3.21.2. PTP Transparent Clocks

A transparent clock is a PTP device that does not process PTP packets but modifies them to account for the residence time correction. That is, the device measures the period that the PTP packets require to pass through the device, and then adds that time span to the PTP packets.

A transparent clock measures the variable delay as the PTP packets pass through a device. The measured delay is calculated by adding the residence time into the correction field of the PTP packet.

A transparent clock can be an end-to-end (E2E) clock or a peer-to-peer (P2P) clock. An E2E transparent clock updates the correction field of the PTP packet only with the residence time. A P2P clock can update the correction field with the residence time of packet and the path delay.

The following limitations apply to a PTP E2E transparent clock on an M4500 series switch:

- You cannot configure a PTP E2E transparent clock at VLAN level. However, the feature functions well for VLAN-tagged packets.

- The PTP E2E transparent clock supports only the following three types of PTP event packets: Sync, Delay_Req, and Delay_Resp. Of these PTP event type packets, the switch can update the PTP packet correction field only for Sync and Delay_Req. A Delay_Resp PTP packet passes through the switch without modification.

- The switch does not detect and modify other PTP event packets, such as Announce, pDelay_Req, pDelay_Resp, and FollowUp.

- The switch supports PTP version 2 (PTPv2) only. PTP version 1 (PTPv1) is not supported.

3.21.3. Manage the PTP End-to-End Transparent Clock

By default, the PTP E2E transparent clock is globally enabled on the switch, that is, on all physical ports. If you configure a LAG, the PTP E2E transparent clock is also enabled on the LAG because the feature is already enabled on all physical ports.

3.21.3.1. Globally Disable PTP End-to-End Transparent Clock

To globally disable PTP End-to-End Transparent Clock:

```
(Netgear Switch) #configure
(Netgear Switch) (Config)#no ptp clock e2e-transparent
```

3.21.3.2. Disable PTP End-to-End Transparent Clock for an Interface

To disable PTP End-to-End Transparent Clock for an interface:

```
(Netgear Switch) #configure
(Netgear Switch) (Config)#interface 0/1
(Netgear Switch) (Interface 0/1)#no ptp clock e2e-transparent
```

3.21.4. Globally Reenable PTP End-to-End Transparent Clock

To globally reenable PTP End-to-End Transparent Clock:

```
(Netgear Switch) #configure
(Netgear Switch) (Config)#ptp clock e2e-transparent
```

3.21.5. Reenable PTP End-to-End Transparent Clock for an Interface

To globally reenable PTP End-to-End Transparent Clock for an interface:

```
(Netgear Switch) #configure
(Netgear Switch) (Config)#interface 0/1
(Netgear Switch) (Interface 0/1)#ptp clock e2e-transparent
```

3.21.6. Display the PTP End-to-End Transparent Clock Status

To display the PTP End-to-End Transparent Clock status:

```
(Netgear Switch) #show ptp clock e2e-transparent


PTP TC global mode.............................. Enabled

Interface   Configured Mode Operational Mode
---------   --------------- ----------------
0/1         Enabled         Disabled
0/2         Enabled         Disabled
0/3         Enabled         Disabled
0/4         Enabled         Disabled
0/5         Enabled         Disabled
0/6         Enabled         Disabled
0/7         Enabled         Disabled
0/8         Enabled         Disabled
0/9         Enabled         Disabled
0/10        Enabled         Disabled
```

```
0/11        Enabled         Disabled
0/12        Enabled         Disabled
0/13        Enabled         Disabled
0/14        Enabled         Disabled
0/15        Enabled         Disabled
0/16        Enabled         Disabled
```

4. Configuring Security Features

4.1. Controlling Management Access

A user can access the switch management interface only after providing a valid user name and password combination that matches the user account information stored in the user database configured on the switch.

The switch supports several features to increase management security and help prevent unauthorized access to the switch configuration interfaces.

4.1.1. Using RADIUS Servers for Management Security

Many networks use a RADIUS server to maintain a centralized user database that contains per-user authentication information. RADIUS servers provide a centralized authentication method for:

- Telnet Access
- Console to Switch Access
- Access Control Port (802.1X)

RADIUS access control utilizes a database of user information on a remote server. Making use of a single database of accessible information—as in an Authentication Server—can greatly simplify the authentication and management of users in a large network. One such type of Authentication Server supports the Remote Authentication Dial In User Service (RADIUS) protocol as defined by RFC 2865.

For authenticating users prior to access, the RADIUS standard has become the protocol of choice by administrators of large accessible networks. To accomplish the authentication in a secure manner, the RADIUS client and RADIUS server must both be configured with the same shared password or secret. This secret is used to generate one-way encrypted authenticators that are present in all RADIUS packets. The secret is never transmitted over the network.

RADIUS conforms to a secure communications client/server model using UDP as a transport protocol. It is extremely flexible, supporting a variety of methods to authenticate and statistically track users. RADIUS is also extensible, allowing for new methods of authentication to be added without disrupting existing functionality.

As a user attempts to connect to the switch management interface, the switch first detects the contact and prompts the user for a name and password. The switch encrypts the supplied information, and a RADIUS client transports the request to a pre-configured RADIUS server.

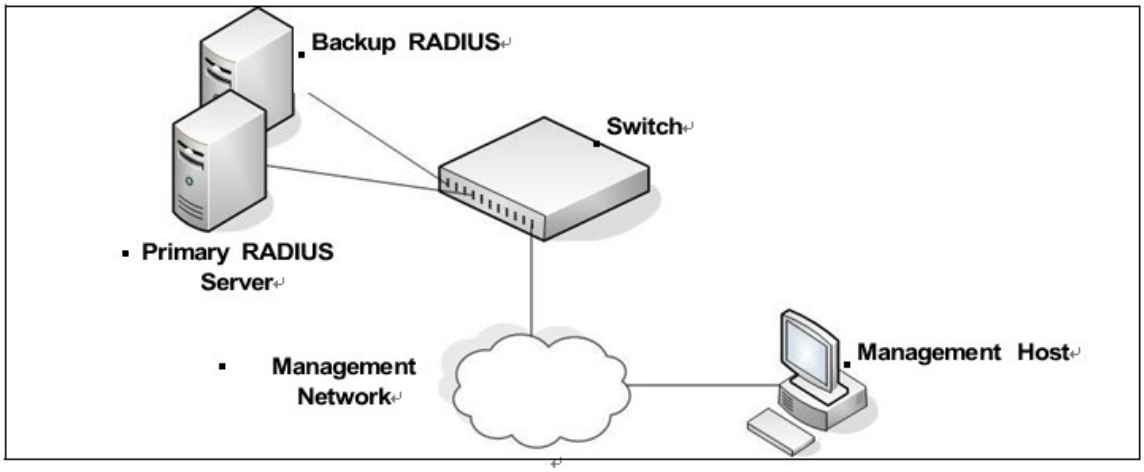


Figure 4-1: RADIUS Topology

The server can authenticate the user itself or make use of a back-end device to ascertain authenticity. In either case a response may or may not be forthcoming to the client. If the server accepts the user, it returns a positive result with attributes containing configuration information. If the server rejects the user, it returns a negative result. If the server rejects the client or the shared secrets differ, the server returns no result. If the server requires additional verification from the user, it returns a challenge, and the request process begins again.

If you use a RADIUS server to authenticate users, you must configure user attributes in the user database on the RADIUS server. The user attributes include the user name, password, and privilege level.

4.1.2. Using TACACS+ to Control Management Access

TACACS+ (Terminal Access Controller Access Control System) provides access control for networked devices via one or more centralized servers. TACACS+ simplifies authentication by making use of a single database that can be shared by many clients on a large network. TACACS+ uses TCP to ensure reliable delivery and a shared key configured on the client and daemon server to encrypt all messages.

If you configure TACACS+ as the authentication method for user login and a user attempts to access the user interface on the switch, the switch prompts for the user login credentials and requests services from the TACACS+ client. The client then uses the configured list of servers for authentication, and provides results back to the switch.

You can configure the TACACS+ server list with one or more hosts defined via their network IP address. You can also assign each a priority to determine the order in which the TACACS+ client will contact them. TACACS+ contacts the server when a connection attempt fails or times out for a higher priority server.

You can configure each server host with a specific connection type, port, timeout, and shared key, or you can use global configuration for the key and timeout.

The TACACS+ server can do the authentication itself, or redirect the request to another back-end device. All sensitive information is encrypted and the shared secret is never passed over the network; it is used only to encrypt the data.

4.1.3. Configuring and Applying Authentication Profiles

A user can access the switch management interface only after providing a valid user name and password combination that matches the user account information stored in the user database configured on the switch.

The switch includes several features to increase management security and help prevent unauthorized access to the CLI.

An authentication profile specifies which authentication method or methods to use to authenticate a user who attempts to access the switch management interface. The profile includes a method list, which defines how authentication is to be performed, and in which order. The list specifies the authentication method to use first, and if the first method returns an error, the next method in the list is tried. This continues until all methods in the list have been attempted. If no method can perform the authentication, then the authentication fails. A method might return an error if, for example, the authentication server is unreachable or misconfigured.

The authentication method can be one or more of the following:

- enable—Uses the enable password for authentication. If there is no enable password defined, then the enable method returns an error.
- line—Uses the Line password for authentication. If there is no line password defined for the access line, then the line method returns an error.
- local— Uses the ID and password in the Local User Database for authentication. If the user ID is not in the local database, access is denied. This method never returns an error. It always permits or denies a user.
- radius—Sends the user's ID and password a RADIUS server to be authenticated. The method returns an error if the switch is unable to contact the server.
- tacacs+— Sends the user's ID and password to a TACACS+ server to be authenticated. The method returns an error if the switch is unable to contact the server.
- none—No authentication is used. This method never returns an error.
- deny—Access is denied. This method never returns an error.

An authentication method might require a user name and password to be supplied, a password only, or no user information. Some methods return errors when authentication fails, while other methods do not. The following table summarizes the method user name/password requirements and error behavior.

Method	*User Name Required*	*Password Required*	*Error Returned*
Local	Yes	Yes	No
RADIUS	Yes	Yes	Yes
TACACS+	Yes	Yes	Yes
Enable	No	Yes	Yes
Line	No	Yes	Yes
None	No	No	No
Deny	No	No	No

Table 4-1: Authentication Method Summary

You can use the same Authentication Profile for all access types, or select or create a variety of profiles based on how a user attempts to access the switch management interface. Profiles can be applied to each of the following access types:

- Login—Authenticates all attempts to login to the switch.
- Enable—Authenticates all attempts to enter Privileged EXEC mode.
- Console—Authenticates access through the console port.
- Telnet—Authenticates users accessing the CLI by using telnet
- SSH—Authenticates users accessing the CLI by using an SSH client.

The following authentication profiles are configured by default:

- defaultList—Method is LOCAL, which means the user credentials are verified against the information in the local user database.
- networkList—Method is LOCAL, which means the user credentials are verified against the information in the local user database.
- enableList—Method is ENABLE, followed by NONE, which means that if the enable password is not configured access is granted. If the enable password is configured and user fails to authenticate then access is not granted.
- enableNetList — Method is ENABLE, followed by DENY, which means that if the enable password is not configured access is denied. This list is applied by default for telnet and SSH. The enable password is not configured by default. That means that, by default, telnet and SSH users will not get access to Privileged EXEC mode. However, a console user always enters the Privileged EXEC mode without entering the enable password in the default configuration.

The methods can be changed, but the preconfigured profiles cannot be deleted or renamed.

4.1.3.1. Configuring Authentication Profiles for Post-based Authentication

In addition to authentication profiles to control access to the management interface, you can configure an authentication profile for IEEE 802.1X port-based access control to control access to the network through the switch ports. To configure a port-based authentication profile, you specify dot1x as the access type, and configure one of the following authentication method: ias, local, none, or radius. The ias method specifies that the 802.1X feature must use the Internal Authentication Server (IAS) database for 801X port-based authentication. The IAS database is stored locally on the switch.

4.1.4. Configuring the Primary and Secondary RADIUS Servers

The commands in this example configure primary and secondary RADIUS servers that the switch will use to authenticate access. The RADIUS servers use the same RADIUS secret.

To configure the switch:

1. Configure the primary and secondary RADIUS servers.

```
(Switch) #configure
(Switch) (Config)#radius server host auth 10.27.65.103
(Switch) (Config)#radius server host auth 10.27.65.114
```

2. Specify which RADIUS server is the primary.

```
(Switch) (Config)#radius server primary 10.27.65.103
(Switch) (Config)#radius server key auth 10.27.65.103
```

3. Configure a shared secret that the switch will use to authenticate with the RADIUS servers.

```
Enter secret (64 characters max):********
Re-enter secret:********
```

4. View the configured RADIUS servers.

```
(Switch) (Config)#exit
(M4500-48XF8C) #show radius servers

Current Host Address           Server Name                      Port  Type      Usage
----    ---------------------- -------------------------------- ----- --------- --------
        10.27.65.114           Default-RADIUS-Server            1812  Secondary Both
  *     10.27.65.103           Default-RADIUS-Server            1812  Primary   Both

 * currently selected server
```

4.1.5. Configuring an Authentication Profile

The commands in this example create a new authentication profile named myList that uses the RADIUS server configured in the previous example to authenticate users who attempt to access the switch management interface by using SSH or Telnet. If the RADIUS authentication is unsuccessful, the switch uses the local user database to attempt to authenticate the users.

To configure the switch:

1. Create an access profile list that uses RADIUS as the first access method and the local user database as the second login method.

```
(Switch) #configure
(Switch) (Config)#aaa authentication login myList radius local
```

Note: The switch attempts to contact the primary RADIUS server that has been configured on the switch. To see an example of how to configure a RADIUS server on the switch, see "Configuring the Primary and Secondary RADIUS Servers."

2. Enter line configuration mode for Telnet and specify that any attempt to access the switch by using Telnet are authenticated using the methods defined in the profile created in the previous step.

```
(Switch) (Config)#line vty
(Switch) (Config-vty)#login authentication myList
(Switch) (Config-vty)#exit
```

3. Enter line configuration mode for SSH and specify that any attempt to access the switch by using SSH are authenticated using the methods defined in the myList profile.

```
(Switch) (Config)#line ssh
(Switch) (Config-ssh)#login authentication myList
(Switch) (Config-ssh)#exit
(Switch) (Config)#exit
```

4. View the current authentication methods and profiles.

```
(Switch) #show authentication methods


Login Authentication Method Lists
---------------------------------
defaultList:  local
networkList:  local
myList:  radiuslocal


Enable Authentication Method Lists
----------------------------------
enableList          :  enable   none
enableNetList       :  enable   deny


 Line     Login Method List   Enable Method List
 -------  -----------------   ------------------
 Console  defaultList         enableList
 Telnet   myList              enableList
 SSH      myList              enableList
```

4.2. Configuring DHCP Snooping, DAI, and IPSG

Dynamic Host Configuration Protocol (DHCP) Snooping, IP Source Guard (IPSG), and Dynamic ARP Inspection (DAI) are layer 2 security features that examine traffic to help prevent accidental and malicious attacks on the switch or network.

DHCP Snooping monitors DHCP messages between a DHCP client and DHCP server to filter harmful DHCP messages and to build a bindings database. The IPSG and DAI features use the DHCP Snooping bindings database to help enforce switch and network security.

IP Source Guard allows the switch to drop incoming packets that do not match a binding in the bindings database. Dynamic ARP Inspection allows the switch to drop ARP packets whose sender MAC address and sender IP address do not match an entry in the DHCP snooping bindings database.

4.2.1. DHCP Snooping Overview

Dynamic Host Configuration Protocol (DHCP) Snooping is a security feature that monitors DHCP messages between a DHCP client and DHCP server to accomplish the following tasks:

- Filter harmful DHCP messages
- Build a bindings database with entries that consist of the following information:

  – MAC address

  – IP address

  – VLAN ID

  – Client port

Entries in the bindings database are considered to be authorized network clients.

DHCP snooping can be enabled on VLANs, and the trust status (trusted or untrusted) is specified on individual physical ports or Port-channels that are members of a VLAN. When a port or Port-channel is configured as untrusted, it could potentially be used to launch a network attack. DHCP servers must be reached through trusted ports.

DHCP snooping enforces the following security rules:

- DHCP packets from a DHCP server (DHCPOFFER, DHCPACK, DHCPNAK, DHCPLEASEQUERY) are dropped if they are received on an untrusted port.
- DHCPRELEASE and DHCPDECLINE messages are dropped if the MAC addresses in the snooping database, but the binding's interface is other than the interface where the message was received.
- On untrusted interfaces, the switch drops DHCP packets with a source MAC address that does not match the client hardware address. This is a configurable option.

4.2.1.1. Populating the DHCP Snooping Bindings Database

The DHCP snooping application uses DHCP messages to build and maintain the binding's database. DHCP snooping creates a tentative binding from DHCP DISCOVER and REQUEST messages. Tentative bindings tie a client to a port (the port where the DHCP client message was received). Tentative bindings are completed when DHCP snooping learns the client's IP address from a DHCP ACK message on a trusted port. DHCP snooping removes bindings in response to DECLINE, RELEASE, and NACK messages. The DHCP snooping application ignores the ACK messages as a reply to the DHCP Inform messages received on trusted ports. You can also enter static bindings into the binding database.

When a switch learns of new bindings or loses bindings, the switch immediately updates the entries in the database. The switch also updates the entries in the binding file. The frequency at which the file is updated is based on a configurable delay, and the updates are batched.

If the absolute lease time of the snooping database entry expires, that entry is removed. Make sure the system time is consistent across the reboots. Otherwise, the snooping entries will not expire properly. If a host sends a DHCP release while the switch is rebooting, when the switch receives the DHCP discovery or request, the client's binding goes to the tentative binding as shown in the following figure.

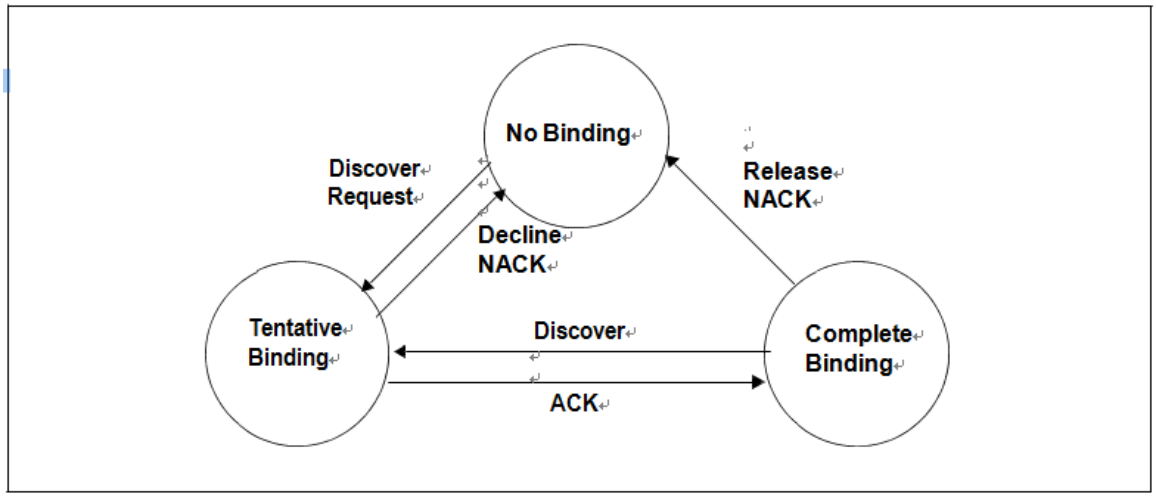


Figure 4-2: DHCP Binding

The binding database includes data for clients only on untrusted ports.

4.2.1.2. DHCP Snooping and VLANs

DHCP snooping forwards valid DHCP client messages received on non-routing VLANs. The message is forwarded on all trusted interfaces in the VLAN.

DHCP snooping can be configured on switching VLANs and routing VLANs. When a DHCP packet is received on a routing VLAN, the DHCP snooping application applies its filtering rules and updates the bindings database. If a client message passes filtering rules, the message is placed into the software forwarding path where it may be processed by the DHCP relay agent, the local DHCP server, or forwarded as an IP packet.

4.2.1.3. DHCP Snooping Logging and Rate Limits

The DHCP snooping application processes incoming DHCP messages. For DHCPRELEASE and DHCPDECLINE messages, the application compares the receive interface and VLAN with the client interface and VLAN in the bindings database. If the interfaces do not match, the application logs the event and drops the message. For valid client messages, DHCP snooping compares the source MAC address to the DHCP client hardware address. When there is a mismatch, DHCP snooping drops the packet and generates a log message if logging of invalid packets is enabled.

If DHCP relay co-exists with DHCP snooping, DHCP client messages are sent to DHCP relay for further processing.

To prevent DHCP packets from being used as a DoS attack when DHCP snooping is enabled, the snooping application enforces a rate limit for DHCP packets received on interfaces. DHCP snooping monitors the receive rate on each interface separately. If the receive rate exceeds a configurable limit, DHCP snooping brings down the interface. Administrative intervention is necessary to enable the port, either by using the **no shutdown** command in Interface Config mode.

4.2.2. IP Source Guard Overview

IPSG is a security feature that filters IP packets based on source ID. This feature helps protect the network from attacks that use IP address spoofing to compromise or overwhelm the network.

The source ID may be either the source IP address or a {source IP address, source MAC address} pair. You can configure:

- Whether enforcement includes the source MAC address
- Static authorized source IDs

The DHCP snooping bindings database and static IPSG entries identify authorized source IDs. IPSG can be enabled on physical and Port-channel ports.

If you enable IPSG on a port where DHCP snooping is disabled or where DHCP snooping is enabled but the port is trusted, all IP traffic received on that port is dropped depending on the admin-configured IPSG entries.

4.2.2.1. IPSG and Port Security

IPSG interacts with port security, also known as port MAC locking to enforce the source MAC address. Port security controls source MAC address learning in the layer 2 forwarding database (MAC address table). When a frame is received with a previously unlearned source MAC address, port security queries the IPSG feature to determine whether the MAC address belongs to a valid binding.

If IPSG is disabled on the ingress port, IPSG replies that the MAC is valid. If IPSG is enabled on the ingress port, IPSG checks the bindings database. If the MAC address is in the bindings database and the binding matches the VLAN the frame was received on, IPSG replies that the MAC is valid. If the MAC is not in the bindings database, IPSG informs port security that the frame is a security violation.

In the case of an IPSG violation, port security takes whatever action it normally takes upon receipt of an unauthorized frame. Port security limits the number of MAC addresses to a configured maximum. If the limit *n* is less than the number of stations *m* in the bindings database, port security allows only *n* stations to use the port. If *n* > *m*, port security allows only the stations in the bindings database.

4.2.3. Dynamic ARP Inspection Overview

Dynamic ARP Inspection (DAI) is a security feature that rejects invalid and malicious ARP packets. DAI prevents a class of man-in-the-middle attacks where an unfriendly station intercepts traffic for other stations by poisoning the ARP caches of its unsuspecting neighbors. The malicious attacker sends ARP requests or responses mapping another station's IP address to its own MAC address.

When DAI is enabled, the switch drops ARP packets whose sender MAC address and sender IP address do not match an entry in the DHCP snooping bindings database. You can optionally configure additional ARP packet validation.

When DAI is enabled on a VLAN, DAI is enabled on the interfaces (physical ports or Port-channels) that are members of that VLAN. Individual interfaces are configured as trusted or untrusted. The trust configuration for DAI is independent of the trust configuration for DHCP snooping.

4.2.3.1. Optional DAI Features

If you configure the MAC address validation option, DAI verifies that the sender MAC address equals the source MAC address in the Ethernet header. There is a configurable option to verify that the target MAC address equals the destination MAC address in the Ethernet header. This check applies only to ARP responses, since the target MAC address is unspecified in ARP requests. You can also enable IP address checking. When this option is enabled, DAI drops ARP packets with an invalid IP address. The following IP addresses are considered invalid:

- 0.0.0.0
- 255.255.255.255
- all IP multicast addresses
- all class E addresses (240.0.0.0/4)
- loopback addresses (in the range 127.0.0.0/8)

The valid IP check is applied only on the sender IP address in ARP packets. In ARP response packets, the check is applied only on the target IP address.

4.2.4. Increasing Security with DHCP Snooping, DAI, and IPSG

DHCP Snooping, IPSG, and DAI are security features that can help protect the switch and the network against various types of accidental or malicious attacks. It might be a good idea to enable these features on ports that provide network access to hosts that are in physically unsecured locations or if network users connect nonstandard hosts to the network.

For example, if an employee unknowingly connects a workstation to the network that has a DHCP server, and the DHCP server is enabled, hosts that attempt to acquire network information from the legitimate network DHCP server might obtain incorrect information from the rogue DHCP server. However, if the workstation with the rogue DHCP server is connected to a port that is configured as untrusted and is a member of a DHCP Snooping-enabled VLAN, the port discards the DHCP server messages.

4.2.5. Configuring DHCP Snooping

In this example, DHCP snooping is enabled on VLAN 100. Ports 1-20 connect end users to the network and are members of VLAN 100. These ports are configured to limit the maximum number of DHCP packets with a rate limit of 100 packets per second. Port-channel 1, which is also a member of VLAN 100 and contains ports 21-24, is the trunk port that connects the switch to the data center, so it is configured as a trusted port.

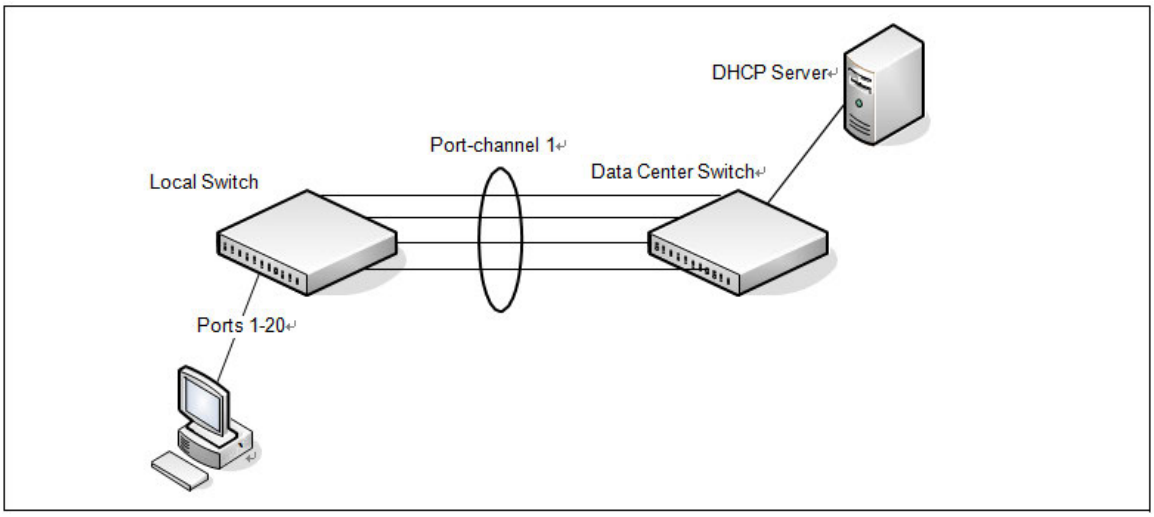


Figure 4-3: DHCP Snooping Configuration Topology

The commands in this example also enforce rate limiting and remote storage of the bindings database. The switch has a limited amount of storage space in NVRAM and flash memory, so you must specify that the DHCP snooping bindings database is stored on an external TFTP server.

To configure the switch:

1. Enable DHCP snooping on VLAN 100.

```
(Switch) #config
(Switch) (Config)#ip dhcp snooping vlan 100
```

2. Configure Port-channel 1, which includes ports 21-24, as a trusted port. All other interfaces are untrusted by default.

```
(Switch) (Config)#interface port-channel 1
(Switch) (if-port-channel ch1)#ip dhcp snooping trust
(Switch) (if-port-channel ch1)#exit
```

3. Enter interface configuration mode for all untrusted interfaces (ports 1-20) and limit the number of DHCP packets that an interface can receive to 100 packets per second. Port-channel 1 is a trusted port and keeps the default value for rate limiting (unlimited).

```
(Switch) (Config)#interface range 0/1-0/20
(Switch) (Interface 0/1-0/20)#ip dhcp snooping limit rate 100
(Switch) (Interface 0/1-0/20)#exit
```

4. Specify that the DHCP snooping database is to be stored remotely in a file called dsDb.txt on a TFTP server with and IP address of 10.131.11.1.

```
(Switch) (Config)#ip dhcp snooping database tftp://10.131.11.1/dsDb.txt
```

5. Enable DHCP snooping for the switch.

```
(Switch) (Config)#ip dhcp snooping
(Switch) (Config)#exit
```

6. View DHCP snooping information.

```
(Switch) #show ip dhcp snooping


DHCP snooping is Enabled
DHCP snooping source MAC verification is enabled
DHCP snooping is enabled on the following VLANs:
100


InterfaceTrustedLog Invalid Pkts
-----------  -------------------------
```

4.2.6. Configuring IPSG

This example builds on the previous example and uses the same topology shown in the previous figure. In this configuration example, IP source guard is enabled on ports 1-20. DHCP snooping must also be enabled on these ports. Additionally, because the ports use IP source guard with source IP and MAC address filtering, port security must be enabled on the ports as well.

To configure the switch:

1. Enter interface configuration mode for the host ports and enable IPSG.

```
(Switch) #config
(Switch) (Config)#interface range 0/1-0/20
(Switch) (Interface 0/1-0/20)#ip verify source port-security
```

2. Enable port security on the ports.

```
(Switch) (Interface 0/1-0/20)#port-security
(Switch) (Interface 0/1-0/20)#exit
(Switch) (Config)#exit
```

3. View IPSG information.

```
(M4500-48XF8C) (Interface 0/1-0/10)#show ip verify source
Interface    Filter Type  IP Address       MAC Address        VLAN
-----------  -----------  ---------------  -----------------  -----
0/1          ip-mac       192.168.3.45     00:1C:23:55:D4:8E  100
0/2          ip-mac       192.168.3.33     00:1C:23:AA:B8:01  100
0/3          ip-mac       192.168.3.18     00:1C:23:55:1B:6E  100
```

4.3. Configuring DHCPv6 Snooping

When enabled on a VLAN, DHCPv6 snooping stands between untrusted ports (those connected to host ports) and trusted ports (those connected to DHCPv6 servers). A VLAN with DHCPv6 snooping enabled forwards DHCPv6 request packets from clients and discards DHCPv6 server reply packets on untrusted ports, and it forwards DHCPv6 server reply packets on trusted ports to DHCPv6 clients.

4.3.1. DHCPv6 Snooping Configuration Example

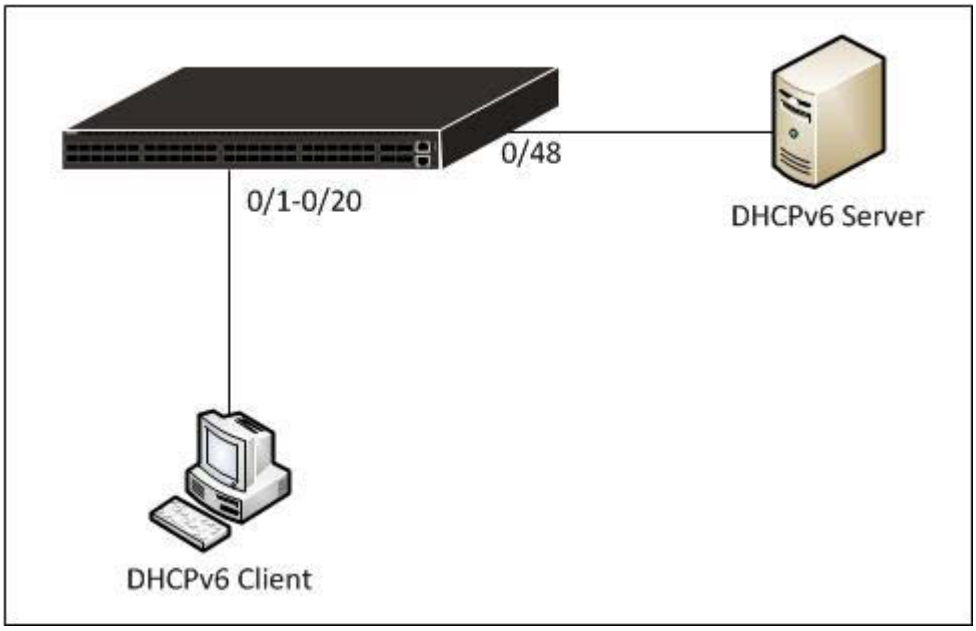


Figure 4-4: DHCPv6 Snooping Configuration Topology

1. Enable DHCPv6 Snooping

```
(Switch) (Config)#ipv6 dhcp snooping
```

2. Enable DHCPv6 Snooping on VLAN 100

```
(Switch) (Config)#ipv6 dhcp snooping vlan 100
```

3. Configure port 48 as a trusted port. All other interfaces are untrusted by default.

```
(Switch) (Config)#
(Switch) (Config)#interface 0/48
(Switch) (Interface 0/48)#ipv6 dhcp snooping trust
(Switch) (Interface 0/48)#exit
```

4. Enter interface configuration mode for all untrusted interfaces (ports 1-20) and limit the number of DHCP packets that an interface can receive to 100 packets per second. Port 48 is a trusted port and keeps the default value for rate limiting (unlimited).

```
(Switch) (Config)#interface range 0/1-0/20
(Switch) (Interface 0/1-0/20)#ipv6 dhcp snooping limit rate 100
(Switch) (Interface 0/1-0/20)#exit
```

5. Verify the configuration.

```
(Switch) #show ipv6 dhcp snooping


DHCP snooping is Enabled
DHCP snooping source MAC verification is enabled
DHCP snooping is enabled on the following VLANs:
100

 Interface    Trusted     Log Invalid Pkts
-----------  ----------   ----------------
0/1          No           No
0/2          No           No
0/3          No           No
0/4          No           No
0/5          No           No
0/6          No           No
0/7          No           No
0/8          No           No
0/9          No           No
0/10         No           No
0/11         No           No
0/12         No           No
0/13         No           No
```

```
0/14          No            No
0/15          No            No
0/16          No            No
0/17          No            No
0/18          No            No
0/19          No            No
0/20          No            No
0/48          Yes           No

(Switch) #show ipv6 dhcp snooping interfaces 0/1

Interface     Trust State     Rate Limit      Burst Interval
                                 (pps)           (seconds)
----------   -------------   -------------   ---------------
0/1          No              100             1



(Switch) #show ipv6 dhcp snooping interfaces 0/48

Interface     Trust State     Rate Limit      Burst Interval
                                 (pps)           (seconds)
----------   -------------   -------------   ---------------
0/48         Yes             None            N/A
```

4.4. ACLs

Access Control Lists (ACLs) are a collection of permit and deny conditions, called rules, which provide security by blocking unauthorized users and allowing authorized users to access specific resources.

ACLs can also provide traffic flow control, restrict contents of routing updates, and decide which types of traffic are forwarded or blocked. ACLs can reside in a firewall router, a router connecting two internal networks, or a Layer 3 switch.

The switch supports ACL configuration in both the ingress and egress direction. Egress ACLs provide the capability to implement security rules on the egress flows (traffic leaving a port) rather than the ingress flows (traffic entering a port). Ingress and egress ACLs can be applied to any physical port, Port-channel, or VLAN routing port.

Depending on whether an ingress or egress ACL is applied to a port, when the traffic enters (ingress) or leaves (egress) a port, the ACL compares the criteria configured in its rules, in order, to the fields in a packet or frame to check for matching conditions. The ACL forwards or blocks the traffic based on the rules.

Note: Every ACL is terminated by an implicit **deny all** rules, which covers any packet not matching a preceding explicit rule.

You can set up ACLs to control traffic at Layer 2, Layer 3, or Layer 4. MAC ACLs operate on Layer 2. IP ACLs operate on Layers 3 and 4. The switch supports both IPv4 and IPv6 ACLs.

4.4.1. MAC ACLs

MAC ACLs are Layer 2 ACLs. You can configure the rules to inspect the following fields of a packet:

- Source MAC address
- Source MAC mask
- Destination MAC address
- Destination MAC mask
- VLAN ID
- Class of Service (CoS) (802.1p)
- EtherType

L2 ACLs can apply to one or more interfaces. Multiple access lists can be applied to a single interface; sequence number determines the order of execution. You can assign packets to queues using the assign queue option.

4.4.2. IP ACLs

IP ACLs classify for Layers 3 and 4 on IPv4 or IPv6 traffic.

Each ACL is a set of up to ten rules applied to inbound traffic. Each rule specifies whether the contents of a given field should be used to permit or deny access to the network, and may apply to one or more of the following fields within a packet:

- Destination IP with wildcard mask
- Destination L4 Port
- Every Packet
- IP DSCP
- IP Precedence
- IP TOS
- Protocol
- Source IP with wildcard mask

- Source L4 port
- IPv4 fragmented packets
- tcp flags
- igmp type
- icmp type
- icmp code
- icmp message

4.4.3. ACL Redirect Function

The redirect function allows traffic that matches a permit rule to be redirected to a specific physical port or Port-channel instead of processed on the original port. The redirect function and mirror function are mutually exclusive. In other words, you cannot configure a given ACL rule with mirror and redirect attributes.

4.4.4. ACL Mirror Function

ACL mirroring provides the ability to mirror traffic that matches a permit rule to a specific physical port or Port-channel. Mirroring is similar to the redirect function, except that in flow-based mirroring a copy of the permitted traffic is delivered to the mirror interface while the packet itself is forwarded normally through the device. You cannot configure a given ACL rule with both mirror and redirect attributes.

Using ACLs to mirror traffic is considered to be flow-based mirroring since the traffic flow is defined by the ACL classification rules. This is in contrast to port mirroring, where all traffic encountered on a specific interface is replicated on another interface.

4.4.5. ACL Logging

ACL Logging provides a means for counting the number of matches against an ACL rule. When you configure ACL Logging, you augment the ACL deny rule specification with a *log* parameter that enables hardware hit count collection and reporting. The switch uses a fixed five minute logging interval, at which time trap log entries are written for each ACL logging rule that accumulated a non-zero hit count during that interval. You cannot configure the logging interval.

4.4.6. Time-based ACLs

The time-based ACL feature allows the switch to dynamically apply an explicit ACL rule within an ACL for a predefined time interval by specifying a time range on a per-rule basis within an ACL, so that the time restrictions are imposed on the ACL rule.

With a time-based ACL, you can define when and for how long an individual rule of an ACL is in effect. To apply a time to an ACL, first you define a specific time interval and then apply it to an individual ACL rule so that it is

operational only during the specified time range, for example, during a specified time period or on specified days of the week.

A time range can be absolute (specific time) or periodic (recurring). If an absolute and periodic time range entry are defined within the same time range, the periodic timer is active only when the absolute timer is active.

Note: Adding a conflicting periodic time range to an absolute time range will cause the time range to become inactive. For example, consider an absolute time range from 8:00 AM Tuesday March 1st

2011 to 10 PM Tuesday March 1st 2011. Adding a periodic entry using the 'weekend' keyword will cause the time-range to become inactive because Tuesdays are not on the weekend.

A named time range can contain up to 10 configured time ranges. Only one absolute time range can be configured per time range. During the ACL configuration, you can associate a configured time range with the ACL to provide additional control over permitting or denying a user access to network resources.

Benefits of using time-based ACLs include:

- Providing more control over permitting or denying a user access to resources, such as an application (identified by an IP address/mask pair and a port number).

- Providing control of logging messages. Individual ACL rules defined within an ACL can be set to log traffic only at certain times of the day so you can simply deny access without needing to analyze many logs generated during peak hours.

4.4.7. ACL Rule Remarks

ACL remarks can be added to ACLs rule to assist users in understanding the rules. Users can add up to 10 remarks per rule, up to 100 characters each (including alphanumeric characters and special characters such as space, hyphen, and underscore. One or more remarks are associated with the rule that is created immediately after the remarks are created and are deleted when the associated rule is deleted. They can be viewed using the **show running-config** command but do not display using the **show access-lists** commands.

4.4.8. ACL Rule Priority

A sequence number cab be added to ACL rule entries to facilitate resequencing them. When a new ACL rule entry is added, a unique sequence number can be specified so that the new ACL rule entry is placed in the desired position in the access list.

If no sequence number is specified, then the rule is assigned a sequence number that is 10 greater than the highest existing sequence number for the rule (that is, it is made the lowest-priority rule); or, if the rule is the first one created for the ACL, it is assigned sequence number 10.

4.4.9. ACL Limitations

The following limitations apply to ingress and egress ACLs.

- Maximum of 100 ACLs.
- Maximum number configurable rules per list is 1023.
- Maximum ACL rules (system-wide) for ingress is 4096
- Maximum ACL rules (system-wide) for egress is 1024
- You can configure mirror or redirect attributes for a given ACL rule, but not both.
- The switch hardware supports a limited number of counter resources, so it may not be possible to log every ACL rule. You can define an ACL with any number of logging rules, but the number of rules that are actually logged cannot be determined until the ACL is applied to an interface. Furthermore, hardware counters that become available after an ACL is applied are not retroactively assigned to rules that were unable to be logged (the ACL must be un-applied then re-applied). Rules that are unable to be logged are still active in the ACL for purposes of permitting or denying a matching packet. If console logging is enabled and the severity is set to Info (6) or a lower severity, a log entry may appear on the screen.
- The order of the rules is important: when a packet matches multiple rules, the first rule takes precedence.

Also, once you define an ACL for a given port, all traffic not specifically permitted by the ACL is denied access.

4.4.10. ACL Configuration Process

To configure ACLs, follow these steps:

1. Create a MAC ACL by specifying a name.
2. Create an IP ACL by specifying a number.
3. Add new rules to the ACL.
4. Configure the match criteria for the rules.
5. Apply the ACL to one or more interfaces.

4.4.11. Preventing False ACL Matches

Be sure to specify ACL access-list, permit, and deny rule criteria as fully as possible to avoid false matches. This is especially important in networks with protocols such as FCoE that have newly-introduced EtherType values. For example, rules that specify a TCP or UDP port value should also specify the TCP or UDP protocol and the IPv4 or IPv6 EtherType. Rules that specify an IP protocol should also specify the EtherType value for the frame.

In general, any rule that specifies matching on an upper-layer protocol field should also include matching constraints for each of the lower-layer protocols. For example, a rule to match packets directed to the well-known UDP port number 22 (SSH) should also include matching constraints on the IP protocol field (protocol=0x11 or UDP) and the EtherType field (EtherType=0x0800 or IPv4). The following table lists commonly-used EtherTypes numbers:

EtherType	*Protocol*
0x0800	Internet Protocol version 4 (IPv4)
0x0806	Address Resolution Protocol (ARP)
0x0842	Wake-on LAN Packet
0x8035	Reverse Address Resolution Protocol (RARP)
0x8100	VLAN tagged frame (IEEE 802.1Q)
0x86DD	Internet Protocol version 6 (IPv6)
0x8808	MAC Control
0x8809	Slow Protocols (IEEE 802.3)
0x8870	Jumbo frames
0x888E	EAP over LAN (EAPOL – 802.1X)
0x88CC	Link Layer Discovery Protocol
0x8906	Fibre Channel over Ethernet
0x8914	FCoE Initialization Protocol
0x9100	Q in Q

Table 4-2: Common EtherType Numbers

The following table lists commonly-used IP protocol numbers:

IP Protocol Number	*Protocol*
0x00	IPv6 Hop-by-hop option
0x01	ICMP
0x02	IGMP
0x06	TCP
0x08	EGP
0x09	IGP
0x11	UDP

Table 4-3: Common IP Protocol Numbers

4.4.12. IPv6 ACL Qualifies

IPv6 ACLs support the following additional qualifiers:

- Qualify fragmented IPv6 packets (packets that have the next header field set to 44).
- Qualify routed IPv6 packets (packets that have a routing extension header (next header field set to 43)).

IP ACLs can be applied on ingress and egress interfaces, VLANs, or both.

4.4.13. ACL Configuration Examples

This section contains the following examples:

- Configuring an IP ACL
- Configuring a MAC ACL
- Configuring a Time-Based ACL

4.4.13.1. Configuring an IP ACL

The commands in this example sets up an IP ACL that permits hosts in the 192.168.77.0/24 subnet to send TCP and UDP traffic only to the host with an IP address of 192.168.77.50. The ACL is applied to port 2 on the switch.

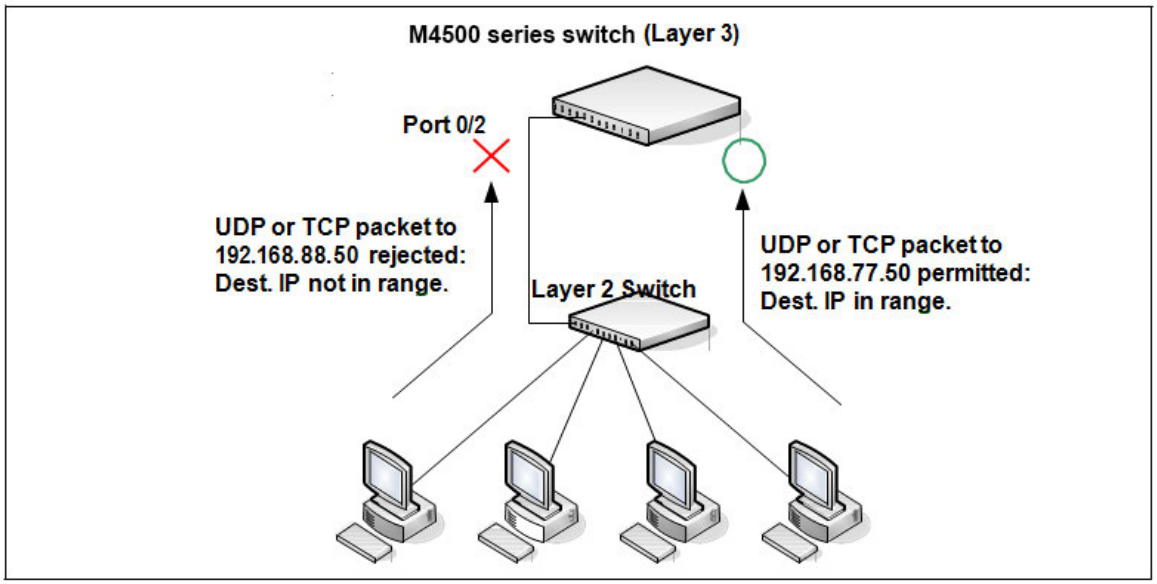


Figure 4-5: IP ACL Example Network Diagram

To configure the switch:

1. Create an extended ACL and configure a rule for the ACL that permits packets carrying TCP traffic that matches the specified Source IP address (192.168.77.0/24), and sends these packets to the specified Destination IP address (192.168.77.50).

```
(Switch) #config
(Switch) (Config)#access-list 100 permit tcp 192.168.77.0 0.0.0.255 192.168.77.50 0.0.0.0
```

2. Define the rule to set similar conditions for UDP traffic as for TCP traffic.

```
(Switch) (Config)#access-list 100 permit udp 192.168.77.0 0.0.0.255 192.168.77.3 0.0.0.255
```

3. Apply the rule to inbound (ingress) traffic on port 2. Only traffic matching the criteria will be accepted on this port.

```
(Switch) (Config)#interface 0/2
(Switch) (Interface 0/2)#ip access-group 100 in
(Switch) (Interface 0/2)#exit
```

4. Verify the configuration.

```
(Switch) #show ip access-lists 100
ACL ID: 100
Inbound Interface(s): 0/2

Sequence Number: 1
Action..................................... permit
Match All.................................. FALSE
Protocol................................... 6(tcp)
Source IP Address.......................... 192.168.77.0
Source IP Wildcard Mask.................... 0.0.0.255
Destination IP Address..................... 192.168.77.50
Destination IP Wildcard Mask............... 0.0.0.0

Sequence Number: 2
Action..................................... permit
Match All.................................. FALSE
Protocol................................... 17(udp)
Source IP Address.......................... 192.168.77.0
Source IP Wildcard Mask.................... 0.0.0.255
Destination IP Address..................... 192.168.77.3
Destination IP Wildcard Mask............... 0.0.0.255
```

4.4.13.2. Configuring a MAC ACL

The following example creates a MAC ACL named mac1 that denies all IPX traffic on all ports. All other types of traffic are permitted.

To configure the switch:

1. Create a MAC Access List named mac1

```
(Switch) #config
(Switch) (Config)#mac access-list extended mac1
```

2. Configure a rule to deny all IPX traffic, regardless of the source or destination MAC address. Before creating the rule, add a remark that identifies the rule.

```
(Switch) (Config-mac-access-list)#remark “Denies all IPX traffic from for any source or dest
MAC”

(Switch) (Config-mac-access-list)#deny any any ipx
```

3. Configure a rule to permit all other types of traffic, regardless of the source or destination MAC address.

```
(Switch) (Config-mac-access-list)#permit any any

(Switch) (Config-mac-access-list)#exit
```

4. Bind the ACL to all ports.

```
(Switch) (Config)#mac access-group mac1 in

(Switch) (Config)#exit
```

5. View information about the configured ACL.

```
(M4500-48XF8C) (Config)#show mac access-lists

Current number of all ACLs: 1  Maximum number of all ACLs: 100


MAC ACL Name                     Rules  Direction  Interface(s)      VLAN(s)
-------------------------------  -----  ---------  ----------------  ----------
mac1                             2      inbound    0/1, 0/2, 0/3,
                                                   0/4, 0/5, 0/6,
                                                   0/7, 0/8, 0/9,
                                                   0/10, 0/11,
                                                   0/12, 0/13,
                                                   0/14, 0/15,
                                                   0/16, 0/17,
                                                   0/18, 0/19,
                                                   0/20, 0/21,
                                                   0/22, 0/23,
                                                   0/24, 0/25,
                                                   0/26, 0/27,
                                                   0/28, 0/29,
                                                   0/30, 0/31,
                                                   0/32, 0/33,
                                                   0/34, 0/35,
                                                   0/36, 0/37,
                                                   0/38, 0/39,
                                                   0/40, 0/41,
                                                   0/42, 0/43,
                                                   0/44, 0/45,
                                                   0/46, 0/47,
                                                   0/48, 0/49,
                                                   0/50, 0/51,
                                                   0/52, 0/53,
                                                   0/54, 0/55,


(Switch) #show mac access-lists mac1


ACL Name: mac1

Inbound Interface(s): 0/1, 0/2, 0/3, 0/4, 0/5, 0/6, 0/7, 0/8, 0/9, 0/10, 0/11, 0/12,
0/13, 0/14,0/15, 0/16, 0/17, 0/18, 0/19, 0/20, 0/21, 0/22, 0/23, 0/24, 0/25, 0/26, 0/27,
0/28, 0/29, 0/30,0/31, 0/32, 0/33, 0/34, 0/35, 0/36, 0/37, 0/38, 0/39, 0/40, 0/41, 0/42,
```

```
0/43, 0/44, 0/45, 0/46,0/47, 0/48, 0/49, 0/50, 0/51, 0/52, 0/53, 0/54, 0/55, 0/56, 0/57,
0/58, 0/59, 0/60, 0/61, 0/62, 0/63, 0/64, 0/65, 0/66, 0/67, 0/68


Sequence Number: 1

Action..................................... deny

Ethertype.................................. ipx


Sequence Number: 2

Action..................................... permit

Match All.................................. TRUE
```

4.4.13.3. Configuring a Time-based ACL

The following example configures an ACL that denies HTTP traffic from 8:00 pm to 12:00 pm and 1:00 pm to

6:00 pm on weekdays and from 8:30 am to 12:30 pm on weekends. The ACL affects all hosts connected to ports that are members of VLAN 100. The ACL permits VLAN 100 members to browse the Internet only during lunch and after hours.

To configure the switch:

1. Create a time range called *work-hours*.

```
(Switch) #config

(Switch) (Config)#time-range work-hours
```

2. Configure an entry for the time range that applies to the morning shift Monday through Friday.

```
(Switch) (Config-time-range)#periodic weekdays 8:00 to 12:00
```

3. Configure an entry for the time range that applies to the afternoon shift Monday through Friday.

```
(Switch) (Config-time-range)#periodic weekdays 13:00 to 18:00
```

4. Configure an entry for the time range that applies to Saturday and Sunday.

```
(Switch) (Config-time-range)#periodic weekend 8:30 to 12:30

(Switch) (Config-time-range)#exit
```

5. Create an extended ACL that denies HTTP traffic during the *work-hours* time range.

```
(Switch) (Config)#access-list 101 deny tcp any any eq http time-range work-hours
```

6. Apply the ACL to ingress traffic in VLAN 100.

```
(Switch) (Config)#ip access-group 101 vlan 100 in

(Switch) (Config)#exit
```

7. Verify the configuration.

```
(Switch) #show ip access-lists 101
```

```
ACL ID: 101
Inbound VLAN ID(s): 100
Sequence Number: 1
Action..................................... deny
Match All.................................. FALSE
Protocol................................... 6(tcp)
Destination L4 Port Keyword................ 80(www/http)
Time Range Name............................ work-hours
Rule Status................................ inactive
```

4.5. Control Plane Policing (CoPP)

Control plane policing (CoPP) uses access control list (ACL) rules to create filters for a system's control plane. That filter prevents traffic not specifically identified as legitimate from reaching the system control plane, rate-limits, traffic to an acceptable level.

CoPP increases security on the system by protecting the routing processor from unnecessary or DoS traffic, giving priority to important control plane and management traffic. CoPP uses a dedicated control plane configuration through the ACL command line interfaces (CLIs) to provide filtering and rate-limiting capabilities for the control plane packets.

The following illustration shows an example how to setting CoPP to deny/permit control packets to switch.

4.5.1. CoPP Configuration Examples

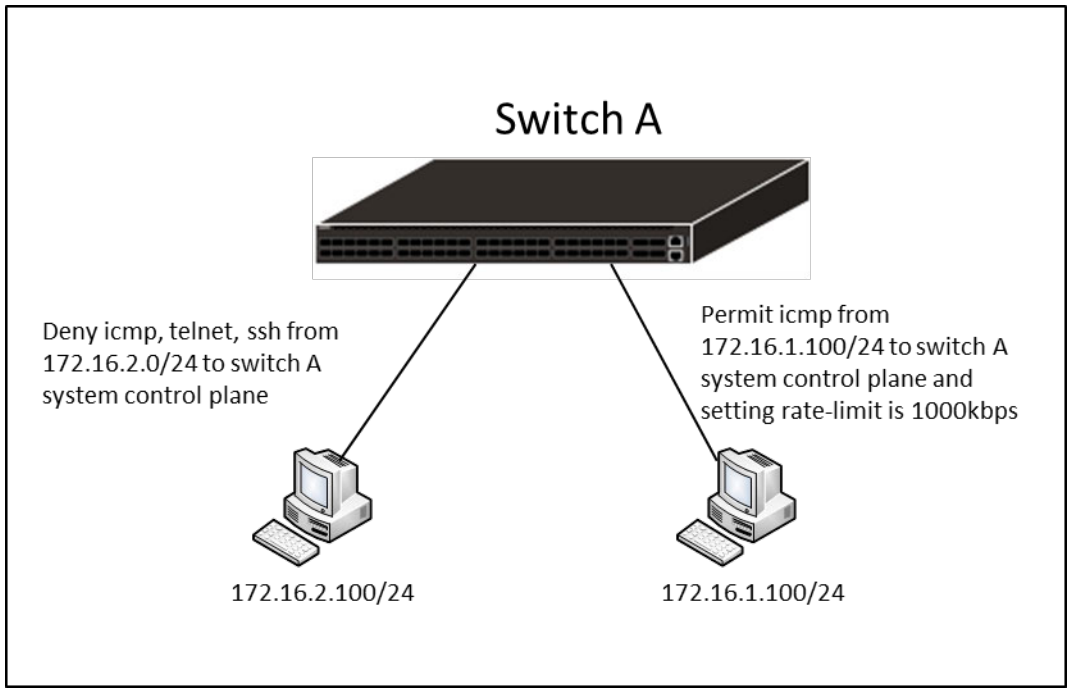


Figure 4-6:CoPP Configuration Topology

1. Create an extended ACL named test and configure rules for the ACL that deny ICMP, Telnet, and SSH packets that match the specified source IP address (172.16.2.100/24).

```
(Switch) (Config)#ip access-list test
(Switch) (Config-ipv4-acl)#deny icmp 172.16.2.100 0.0.0.255 any
(Switch) (Config-ipv4-acl)#deny tcp 172.16.2.100 0.0.0.255 any eq 22
(Switch) (Config-ipv4-acl)#deny tcp 172.16.2.100 0.0.0.255 any eq telnet
```

2. Permit ICMP packets for the specified source IP address (172.16.1.100/24) and configure a rate limit of 1000 kbps.

```
(Switch) (Config-ipv4-acl)#permit icmp 172.16.1.100 0.0.0.255 any rate-limit 1000 1
```

3. Permit any other packets.

```
(Switch) (Config-ipv4-acl)#permit every
(Switch) (Config-ipv4-acl)#exit
```

4. Binding ACL to control plane.

```
(Switch) (Config)#interface control-plane
(Switch) (if-control-plane)#ip access-group test
```

5. Verify the configuration.

```
(Switch) #show ip access-lists test

ACL Name: test
Outbound Interface(s): control-plane



Sequence Number: 1
Action..................................... deny
Match All.................................. False
IPv4 Protocol.............................. 1(icmp)
Source IP Address.......................... 172.16.2.100
Source IP Wildcard Mask.................... 0.0.0.255

Sequence Number: 2
Action..................................... deny
Match All.................................. False
IPv4 Protocol.............................. 6(tcp)
```

```
Source IP Address.............................. 172.16.2.100
Source IP Wildcard Mask........................ 0.0.0.255
Destination L4 Port Keyword.................... 22

Sequence Number: 3
Action......................................... deny
Match All...................................... False
IPv4 Protocol.................................. 6(tcp)
Source IP Address.............................. 172.16.2.100
Source IP Wildcard Mask........................ 0.0.0.255
Destination L4 Port Keyword.................... 23(telnet)

Sequence Number: 4
Action......................................... permit
Match All...................................... False
IPv4 Protocol.................................. 1(icmp)
Source IP Address.............................. 172.16.1.100
Source IP Wildcard Mask........................ 0.0.0.255
Committed Rate................................. 1000
Committed Burst Size........................... 1

Sequence Number: 5
Action......................................... permit
Match All...................................... TRUE
```

5. Configuring Quality of Service

5.1. CoS

The CoS feature lets you give preferential treatment to certain types of traffic over others. To set up this preferential treatment, you can configure the ingress ports, the egress ports, and individual queues on the egress ports to provide customization that suits your environment.

The level of service is determined by the egress port queue to which the traffic is assigned. When traffic is queued for transmission, the rate at which it is serviced depends on how the queue is configured and possibly the amount of traffic present in other queues for that port. Some traffic is classified for service (i.e., packet marking) before it arrives at the switch. If you decide to use these classifications, you can map this traffic to egress queues by setting up a CoS Mapping table.

Each ingress port on the switch has a default priority value (set by configuring VLAN Port Priority in the Switching sub-menu) that determines the egress queue its traffic gets forwarded to. Packets that arrive without a priority designation, or packets from ports you've identified as "untrusted," get forwarded according to this default.

5.1.1. Trusted and Untrusted Port Modes

Ports can be configured in *trusted* mode or *untrusted* mode with respect to ingress traffic.

Ports in Trusted Mode: When a port is configured in trusted mode, the system accepts at face value a priority designation encoded within packets arriving on the port. You can configure ports to trust priority designations based on one of the following fields in the packet header:

- 802.1 Priority: values 0–7
- IP DSCP: values 0–63

A mapping table associates the designated field values in the incoming packet headers with a traffic class priority (actually a CoS traffic queue).

Ports in Untrusted Mode: If you configure an ingress port in untrusted mode, the system ignores any priority designations encoded in incoming packets, and instead sends the packets to a traffic queue based on the ingress port's default priority.

5.1.2. Traffic Shaping on Egress Traffic

For slot/port interfaces, you can specify a traffic shaping rate for the port (in Kbps) for egress traffic. The traffic shaping rate specifies an upper limit of the transmission bandwidth used.

5.1.3. Defining Traffic Queues

For each queue, you can specify:

- Minimum bandwidth guarantee: A percentage of the port's maximum negotiated bandwidth reserved for the queue.

- Scheduler type – strict/weighted:

  – S t r i c t priority scheduling gives an absolute priority, with traffic in the highest priority queues always sent first, and traffic in the lowest priority queues always sent last.

  – W e i g h t e d scheduling requires a specification of priority for each queue relative to the other queues, based on their minimum bandwidth values.

5.1.3.1. Supported Queue Management Methods

The switch supports the following methods, configurable per-interface-queue, for determining which packets are dropped when the queue is full:

- Taildrop: Any packet forwarded to a full queue is dropped regardless of its importance.

- Weighted Random Early Detection (WRED) drops packets selectively based their drop precedence level.

For each of four drop precedence levels on each WRED-enabled interface queue, you can configure the following parameters:

– Minimum Threshold: A percentage of the total queue size below which no packets of the selected drop precedence level are dropped.

– Maximum Threshold: A percentage of the total queue size above which all packets of the selected drop precedence level are dropped.

– Drop Probability: When the queue depth is between the minimum and maximum thresholds, this value provides a scaling factor for increasing the number of packets of the selected drop precedence level that are dropped as the queue depth increases.

5.1.3.2. CoS Configuration Example

The following figure illustrates the network operation as it relates to CoS mapping and queue configuration.

Four packets arrive at the ingress port 0/10 in the order A, B, C, and D. Port 0/10 is configured to trust the 802.1p field of the packet, which serves to direct packets A, B, and D to their respective queues on the egress port. These three packets utilize the 802.1p to CoS Mapping Table for port 0/10.

In this example, the 802.1p user priority 3 is configured to send the packet to queue 5 instead of the default queue 3. Since packet C does not contain a VLAN tag, the 802.1p user priority does not exist, so Port 0/10 relies on its default port priority (2) to direct packet C to egress queue 1.

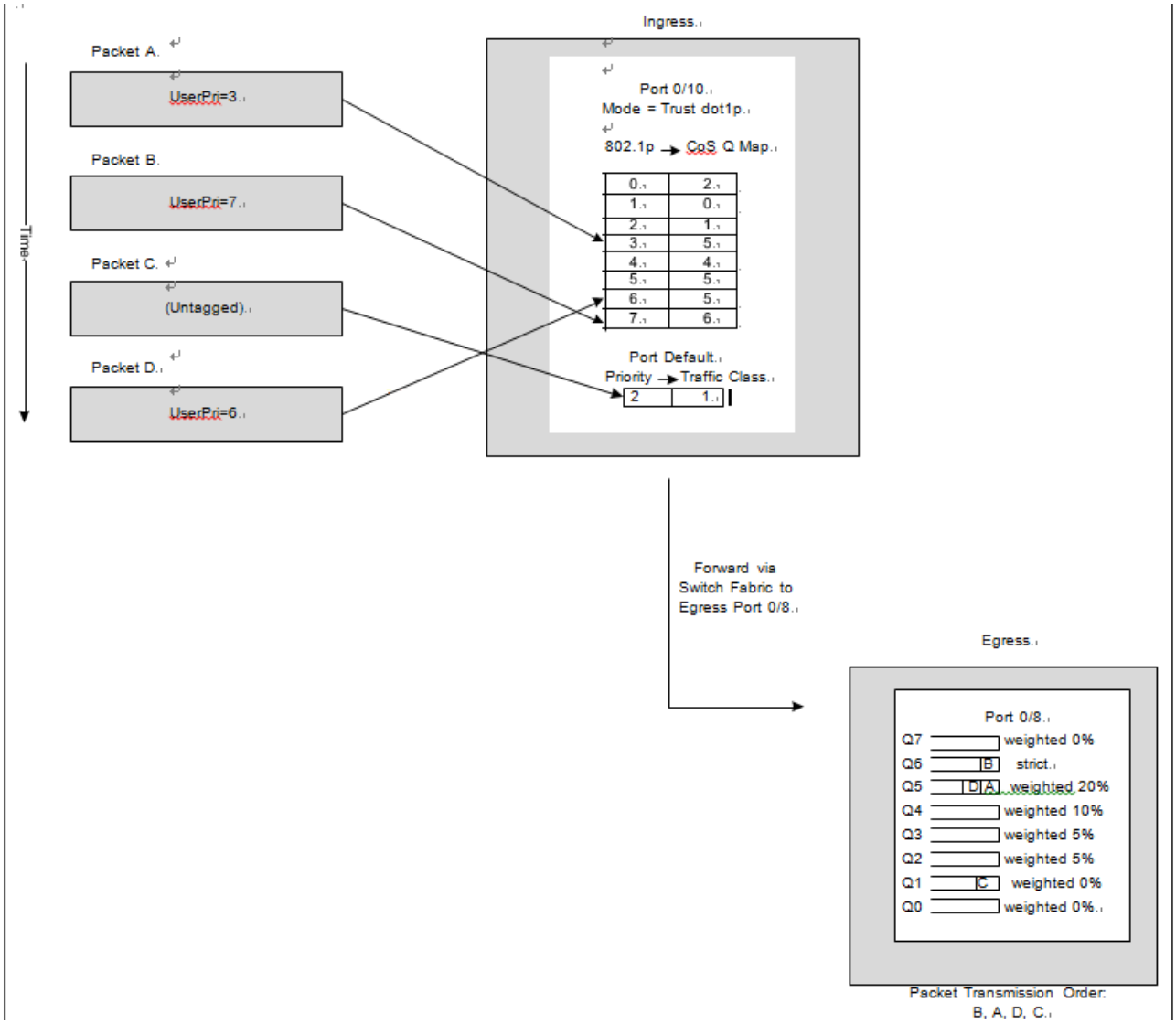


Figure 5-1: CoS Mapping and Queue Configuration

Continuing this example, the egress port 0/8 is configured for strict priority on queue 6, and a weighted scheduling scheme is configured for queues 5-0. Assuming queue 5 has a higher weighting than queue 1 (relative weight values shown as a percentage, with 0% indicating the bandwidth is not guaranteed), the queue service order is 6 followed by 5 followed by 1. Assuming each queue unloads all packets shown in the diagram, the packet transmission order as seen on the network leading out of Port 0/8 is B, A, D, C. Thus, packet B, with its higher user precedence than the others, is able to work its way through the device with minimal delay and is transmitted ahead of the other packets at the egress port.

The following commands configure port 10 (ingress interface) and Port 8 (egress interface).

1. Configure the Trust mode for port 10.

```
(Switch) #config
(Switch) (Config)#interface 0/10
(Switch) (Interface 0/10)#queue trust dot1p
```

2. For port 10, configure the 802.1p user priority 3 to send the packet to queue 5 instead of the default queue (queue 3).

```
(Switch) (Interface 0/10)#queue cos-map 3 5
```

3. For port 10, specify that untagged VLAN packets should have a default priority of 2.

```
(Switch) (Interface 0/10)#switchport priority 2
(Switch) (Interface 0/10)#exit
```

4. For Port 8, the egress port, configure a weighted scheduling scheme for queues 5–0.

```
(Switch) (Config)#interface 0/8
(Switch) (Interface 0/8)#queue cos-queue min-bandwidth 0 0 5 5 10 20 40 0
```

5. Configure Port 8 to have strict priority on queue 6.

```
(Switch) (Interface 0/8)#queue cos-queue strict 6
```

6. View the configuration.

```
(M4500-48XF8C) #show queue cos-queue 0/8

Interface..................................... 0/8
Interface Shaping Rate........................ 0
WRED Decay Exponent........................... 9

Queue ID   Min. Bandwidth   Scheduler Type   Queue Management Type
--------   --------------   --------------   ---------------------
0          0                Weighted         Tail Drop
1          0                Weighted         Tail Drop
2          5                Weighted         Tail Drop
3          5                Weighted         Tail Drop
4          10               Weighted         Tail Drop
5          20               Weighted         Tail Drop
6          40               Strict           Tail Drop
7          0                Weighted         Tail Drop
```

5.2. DiffServ

Standard IP-based networks are designed to provide *best effort* data delivery service. Best effort service implies that the network delivers the data in a timely fashion, although there is no guarantee that it will. During times of congestion, packets may be delayed, sent sporadically, or dropped. For typical Internet applications, such as email and file transfer, a slight degradation in service is acceptable and in many cases unnoticeable.

Conversely, any degradation of service has undesirable effects on applications with strict timing requirements, such as voice or multimedia.

5.2.1. DiffServ Functionality and Switch Roles

How you configure DiffServ support varies depending on the role of the switch in your network:

- **Edge device**: An edge device handles ingress traffic, flowing towards the core of the network, and egress traffic, flowing away from the core. An edge device segregates inbound traffic into a small set of traffic classes, and is responsible for determining a packet's classification. Classification is primarily based on the contents of the Layer 3 and Layer 4 headers, and is recorded in the Differentiated Services Code Point (DSCP) added to a packet's IP header.

- **Interior node**: A switch in the core of the network is responsible for forwarding packets, rather than for classifying them. It decodes the DSCP in an incoming packet, and provides buffering and forwarding services using the appropriate queue management algorithms.

Before configuring DiffServ on the switch, you must determine the QoS requirements for the network as a whole. The requirements are expressed in terms of rules, which are used to classify inbound or outbound traffic on a particular interface.

5.2.2. Elements of DiffServ Configuration

During configuration, you define DiffServ rules in terms of classes, policies, and services:

- **Class**: A class consists of a set of rules that identify which packets belong to the class. Inbound traffic is separated into traffic classes based on Layer 2, Layer 3, and Layer 4 header data. The class type **All** is supported; this specifies that every match criterion defined for the class must be true for a match to occur.

- **Policy**: A policy defines the QoS attributes for one or more traffic classes. An attribute identifies the action taken when a packet matches a class rule. An example of an attribute is to mark a packet. The switch supports the ability to assign traffic classes to output CoS queues, and to mirror incoming packets in a traffic stream to a specific egress interface (physical port or Port-channel).

The switch supports the **Traffic Conditioning Policy** type which is associated with an inbound traffic class and specifies the actions to be performed on packets meeting the class rules:

  – Marking the packet with a given DSCP, IP precedence, or CoS value. Traffic to be processed by the

  DiffServ feature requires an IP header if the system uses IP Precedence or IP DSCP marking.

  – Policing packets by dropping or re-marking those that exceed the class's assigned data rate.

  – Counting the traffic within the class.

- **Service:** Assigns a policy to an interface for inbound traffic.

5.2.3. Configuration DiffServ to Provide Subnets Equal Access to External Network

This example shows how a network administrator can provide equal access to the Internet (or other external network) to different departments within a company. Each of four departments has its own Class B subnet that is allocated 25% of the available bandwidth on the port accessing the Internet.

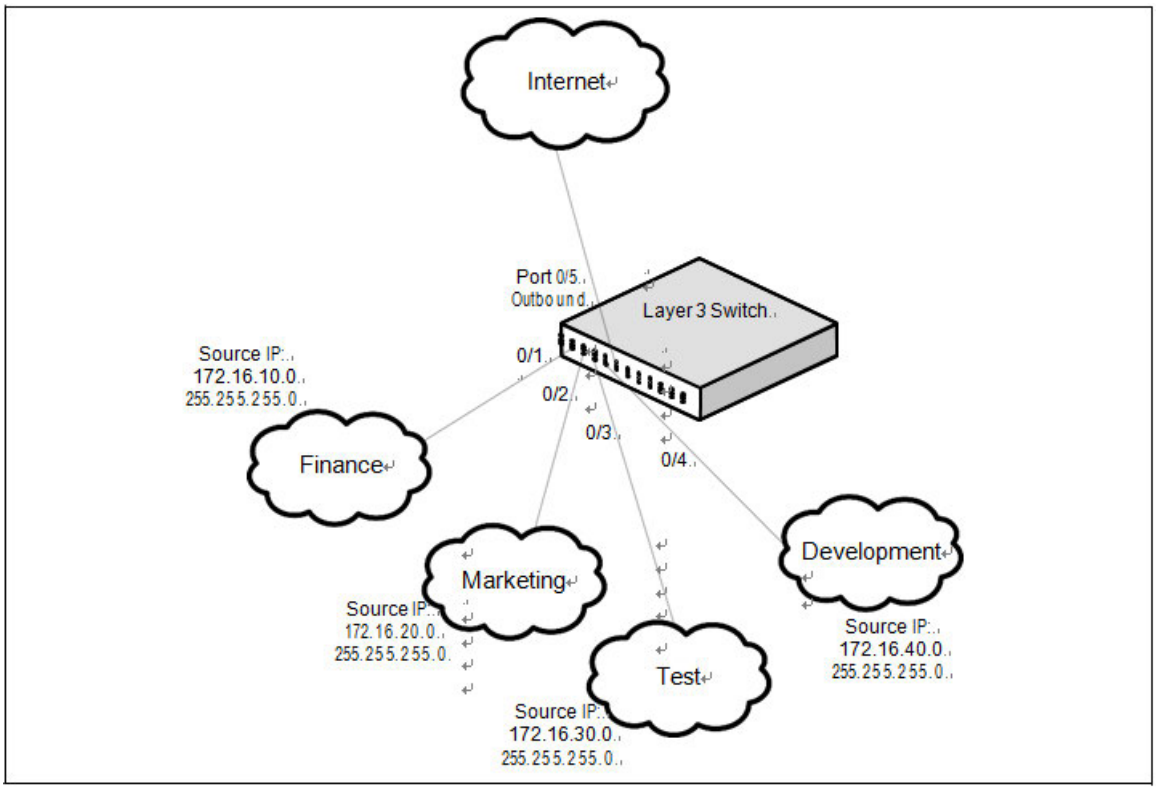


Figure 5-2: DiffServ Internet Access Example Network Diagram

The following commands show how to configure the DiffServ example depicted in the previous figure.

1. Enable DiffServ operation for the switch.

```
(Switch) #config
(Switch) (Config)#diffserv
```

2. Create a DiffServ class of type *all* for each of the departments, and name them. Also, define the match criteria—Source IP address—for the new classes.

```
(Switch) (Config)#class-map match-all finance_dept
(Switch) (Config-classmap)#match srcip 172.16.10.0 255.255.255.0
(Switch) (Config-classmap)#exit
(Switch) (Config)#class-map match-all marketing_dept
(Switch) (Config-classmap)#match srcip 172.16.20.0 255.255.255.0
(Switch) (Config-classmap)#exit
(Switch) (Config)#class-map match-all test_dept
(Switch) (Config-classmap)#match srcip 172.16.30.0 255.255.255.0
```

```
(Switch) (Config-classmap)#exit
(Switch) (Config)#class-map match-all development_dept
(Switch) (Config-classmap)#match srcip 172.16.40.0 255.255.255.0
(Switch) (Config-classmap)#exit
```

3. Create a DiffServ policy for inbound traffic named *internet_access,* adding the previously created department classes as instances within this policy. This policy uses the assign-queue attribute to put each department's traffic on a different egress queue. This is how the DiffServ inbound policy connects to the CoS queue settings established below.

```
(Switch) (Config)#policy-map internet_access in
(Switch) (Config-policy-map)#class finance_dept
(Switch) (Config-policy-classmap)#assign-queue 1
(Switch) (Config-policy-classmap)#exit
(Switch) (Config-policy-map)#class marketing_dept
(Switch) (Config-policy-classmap)#assign-queue 2
(Switch) (Config-policy-classmap)#exit
(Switch) (Config-policy-map)#class test_dept
(Switch) (Config-policy-classmap)#assign-queue 3
(Switch) (Config-policy-classmap)#exit
(Switch) (Config-policy-map)#class development_dept
(Switch) (Config-policy-classmap)#assign-queue 4
(Switch) (Config-policy-classmap)#exit
(Switch) (Config-policy-map)#exit
```

4. Attach the defined policy to interfaces 0/1 through 0/4 in the inbound direction

```
(Switch) (Config)#interface range 0/1-0/4
(Switch) (Interface 0/1-0/4)#service-policy in internet_access
(Switch) (Interface 0/1-0/4)#exit
```

5. Set the CoS queue configuration for the (presumed) egress interface 0/1 such that each of queues 1, 2, 3 and 4 get a minimum guaranteed bandwidth of 25%. All queues for this interface use weighted round robin scheduling by default. The DiffServ inbound policy designates that these queues are to be used for the departmental traffic through the assign-queue attribute. It is presumed that the switch will forward this traffic to interface 0/1 based on a normal destination address lookup for internet traffic.

```
(Switch) (Config)#interface 0/5
(Switch) (Interface 0/5)#queue cos-queue min-bandwidth 0 25 25 25 25 0 0 0
(Switch) (Interface 0/5)#exit
(Switch) (Config)#exit
```

6. Configuring Switch Management Features

6.1. Managing Images and Files

Switches maintain several different types of files on the flash file system. The following table describes the files that you can manage. You use the copy command to copy a source file to a destination file. The copy command may permit the following actions (depending on the file type):

- Copy a file from the switch to a remote server
- Copy a file from a remote server to the switch
- Overwrite the contents of the destination file with the contents of the source file.

File	Description
startup-config	Contains the software configuration that loads during the boot process.
running-config	Contains the current switch configuration.
factory-defaults	Contains the software configuration that can be uses to load during the boot process or after clearing the configuration.
backup-config	An additional configuration file that serves as a backup. You can copy the starting-config file to the backup-config file.
script	Text file with CLI commands. When you apply a script on the switch, the commands are executed and added to the running-config.
CLI Banner	Text file containing the message that displays upon connecting to the switch or logging on to the switch by using the CLI.
Log files	Trap, error, or other log files that provides various information about events that occur on the switch.
SSH key files	Contains information to authenticate SSH sessions. The switch supports the following files for SSH: • SSH-1 RSA Key File • SSH-2 RSA Key File (PEM Encoded) • SSH-2 Digital Signature Algorithm (DSA) Key File (PEM Encoded) • SSH user Public Key file for current user. It supports DSA or RSA Key file of OpenSSH key format. ***Note:*** If you use the CLI to manage the switch over an SSH connection, you must copy the appropriate key files to the switch.

Table 6-1: Files to Manage

6.1.1. Supported File Management Methods

For most file types, you can use any of the following protocols to download files from a remote system to the switch or to upload files from the switch to a remote system:

- FTP
- TFTP
- SFTP
- SCP

6.1.2. Uploading and Downloading Files

To use FTP, TFTP, SFTP, or SCP for file management, you must provide the IP address of the remote system that is running the appropriate server (FTP, TFTP, SFTP, or SCP). Make sure there is a route from the switch to the remote system. You can use the `ping` command from the CLI to verify that a route exists between the switch and the remote system.

If you are copying a file from the remote system to the switch, be sure to provide the correct path to the file (if the file is **not** in the root directory) and the correct file name.

6.1.3. Managing Configuration Files

Configuration files contain the CLI commands that change the switch from its default configuration. The switch can maintain three separate configuration files: startup-config, running-config, and backup-config. The switch loads the startup-config file when the switch boots. Any configuration changes that take place after the boot process completes are written to the running-config file. The backup-config file does not exist until you explicitly create one by copying an existing configuration file to the backup-config file or downloading a backup-config file to the switch.

You can also create configuration scripts, which are text files that contains CLI commands.

When you apply (run) a configuration script on the switch, the commands in the script are executed in the order in which they are written as if you were typing them into the CLI. The commands that are executed in the configuration script are added to the running-config file.

You might upload a configuration file from the switch to a remote server for the following reasons:

- To create a backup copy
- To use the configuration file on another switch
- To manually edit the file

You might download a configuration file from a remote server to the switch for the following reasons:

- To restore a previous configuration

- To load the configuration copied from another switch
- To load the same configuration file on multiple switches

Use a text editor to open a configuration file and view or change its contents.

6.1.3.1. Editing and Downloading Configuration Files

Each configuration file contains a list of executable CLI commands. The commands must be complete and in a logical order, as if you were entering them by using the switch CLI.

When you download a startup-config or backup-config file to the switch, the new file replaces the previous version. To change the running-config file, you execute CLI commands either by typing them into the CLI or by applying a configuration script with the `script apply` command.

6.1.3.2. Creating and Applying Configuration Scripts

When you use configuration scripting, keep the following considerations and rules in mind:

- The application of scripts is partial if the script fails. For example, if the script executes four of ten commands and the script fails, the script stops at four, and the final six commands are not executed.
- Scripts cannot be modified or deleted while being applied.
- Validation of scripts checks for syntax errors only. It does not validate that the script will run.
- The file extension must be `.scr`.
- There is no limit on the maximum number of scripts files that can be stored on the switch within a given storage space limit.
- The combined size of all script files on the switch cannot exceed 2048 Kbytes. The zlib compression technique is applied to script files to decrease script file size.

You can type single-line annotations in the configuration file to improve script readability. The exclamation point (!) character flags the beginning of a comment. Any line in the file that begins with the "!" character is recognized as a comment line and ignored by the parser. Do not use a comment character anywhere in a line that contains a command.

The following example shows annotations within a file (commands are bold):

```
!Configuration script for mapping lab hosts to IP addresses
!Enter Global Config mode and map host name to address configure
ip host labpc1 192.168.3.56
ip host labpc2 192.168.3.57 ip host labpc3 192.168.3.58 exit
! End of the script file
```

6.1.3.3. Non-Disruptive Configuration Management

The Non-Disruptive Configuration feature can apply a new configuration file without disrupting the operation of features that are unchanged by the new configuration.

In the datacenter network, where the network administrator may manage thousands of switches, when the switch configuration is changed by uploading a new configuration file to it, the switch can gracefully resolve any differences between the running configuration and the new configuration. For example, if the switch has VLANs

10, 20, and 30 configured, and the new configuration has VLANs 10, 20, and 40, the switch deletes VLAN 30 and creates VLAN 40 without disturbing traffic forwarding on VLANs 10 and 20.

Without this feature, to upgrade to a new configuration, you must either provide a new configuration file and restart the switch or upload a 'delta' configuration. Restarting the switch is disruptive, and managing delta configurations is difficult on a large scale.

The following commands can be used to apply the configuration gracefully.

- `reload configuration` — Applies the startup-config gracefully.
- `reload configuration <scriptfile>` — Applies the given script file gracefully.

6.1.4. Saving the Running Configuration

Changes you make to the switch configuration while the switch is operating are written to the running-config. These changes are not automatically written to the startup-config. When you reload the switch, the startup-config file is loaded. If you reload the switch (or if the switch resets unexpectedly), any settings in the running-config that were not explicitly saved to the startup-config are lost. You must save the running-config to the startup-config to ensure that the settings you configure on the switch are saved across a switch reset.

To save the running-config to the startup-config from the CLI, use the `copy running-config startup-config` command.

6.1.5. File and Image Management Configuration Examples

This section contains the following examples:

- Upgrading the Firmware
- Managing Configuration Scripts

6.1.5.1. Upgrading the Firmware

This example shows how to download an ONIE installer file or a QNOS image file to the switch and install the file.

- File name for the switch image + Linux OS:
  onie-installer-x86_64-netgear_m4500-48xf8c_dnv-7.0.0.8

- File name for the switch image only:
  QNOS-m4500-48xf8c-7.0.0.8.deb

Use the following steps to prepare the download, and then download and upgrade the switch image.

1. Connect to the switch using the console or SSH method, as described in the “Login User ID and Password” section.

2. After you log in to the switch CLI, press **<Ctrl> + z** or enter **logout** at the switch CLI prompt to exit the switch CLI and display the following menu.

```
====================================
NETGEAR M4500 Menu
====================================
1: CLI Console
2: Firmware update using SCP
3: Firmware update using TFTP
4: Reboot
====================================

Enter your menu option:
```

3. Select one of the following options:

- Option 2. Select this option to download the image using SCP.
- Option 3. Select this option to download image using TFTP.

4. Provide the server and image details, depending on your selection form the menu.

5. If you downloaded and installed the ONIE image in the previous step, reboot the switch using option 4. The switch enters ONIE install mode and installs the ONIE image. When you see the login prompt, log in and check the software version.

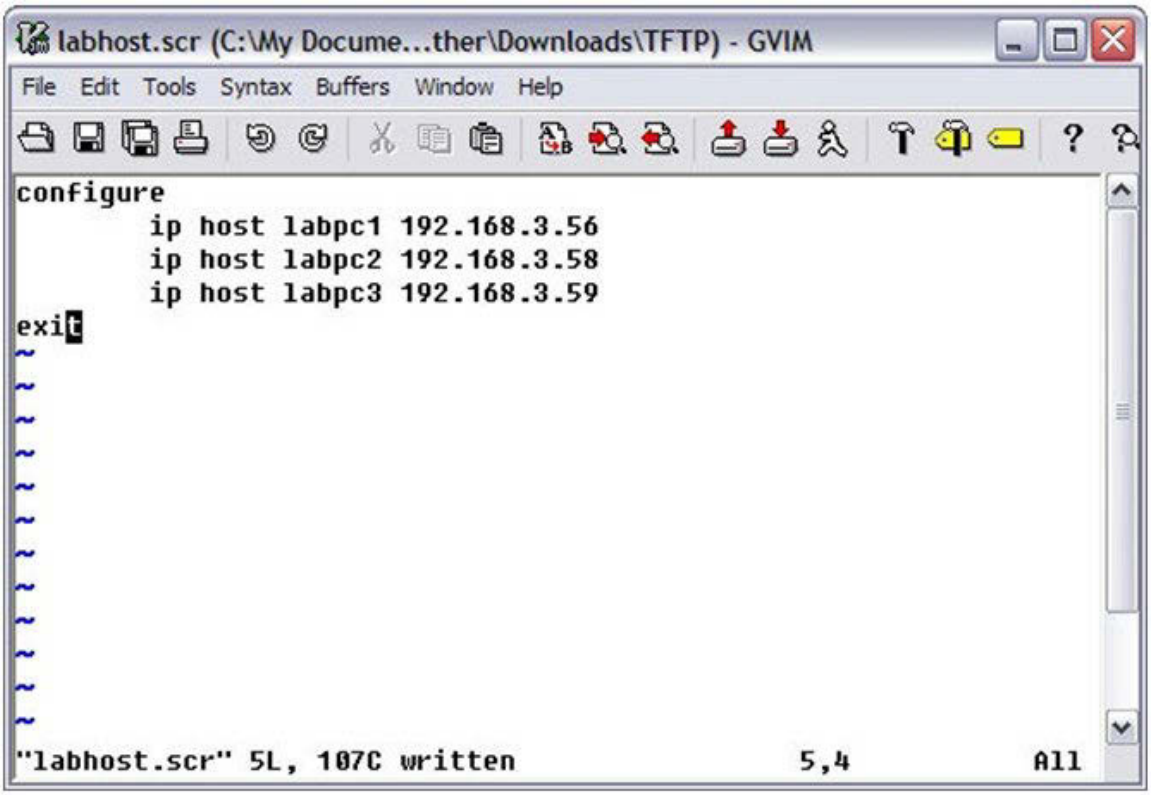
6. If you downloaded and installed the QNOS image in the previous step, no reboot is required. Select option 1 to enter the switch CLI and check the software version.

6.1.5.2. Managing Configuration Scripts

This example shows how to create a configuration script that adds three host name-to-IP address mappings to the host table.

To configure the switch:

1. Open a text editor on an administrative computer and type the commands as if you were entering them by using the CLI.



2. Save the file with an *.scr extension and copy it to the appropriate directory on your TFTP server.

3. Download the file from the TFTP server to the switch.

```
(Switch) #copy tftp://172.16.1.102/labhost.scr script labhost.scr


Mode........................................... TFTP

Set Server IP.................................. 172.16.1.102

Path........................................... ./
Filename....................................... labhost.scr Data
Type........................................... Config Script Destination
Filename....................................... labhost.scr


Management access will be blocked for the duration of the transfer

Are you sure you want to start? (y/n)
```

4. After you confirm the download information and the script successfully downloads, it is automatically validated for correct syntax.

```
Are you sure you want to start? (y/n) y

Validating configuration script...

Configure

ip host labpc1 192.168.3.56 ip host labpc2 192.168.3.58 ip host labpc3 192.168.3.59

exit
```

```
Configuration script validated.
File transfer operation completed successfully.
```

5. Run the script to execute the commands.

```
(Switch) #script apply labhost.scr
Are you sure you want to apply the configuration script? (y/n)y configure
ip host labpc1 192.168.3.56
ip host labpc2 192.168.3.58 ip host labpc3 192.168.3.59
exit
Configuration script 'labhost.scr' applied.
```

6. Verify that the script was successfully applied.

```
(M4500-48XF8C) #show hosts


Host name..................................... M4500-48XF8C
Default domain................................ Domain name is not configured
Default domain list........................... Domain Name List is not configured
Domain Name Lookup............................ Enabled
Number of retries............................. 2
Retry timeout period.......................... 3
Name servers (Preference order)............... 10.1.1.7, 10.1.1.6

Dns Client Source Interface................... (not configured)

Configured host name-to-address mapping:

 Host                                Addresses
------------------------ ----------------------
No host name is configured to IP address

 Host                   Total   Elapsed    Type        Addresses
---------------------- -------  -------    ----        --------------
No hostname is mapped to an IP address
```

6.2. Enabling Automatic System Configuration

The Auto Install feature can automatically obtain configuration information when the switch boots. Auto Install begins the automatic download and installation process when the switch boots and loads a saved configuration that has the persistent Auto Install mode enabled. Additionally, the switch supports a non-persistent Auto Install mode so that Auto Install can be stopped or restarted at any time during switch operation.

6.2.1. DHCP Auto Install Process

The switch can use a DHCP server to obtain configuration information from a TFTP server. DHCP Auto Install is accomplished in three phases:

1. Assignment or configuration of an IP address for the switch

2. Assignment of a TFTP server

3. Obtaining a configuration file for the switch from the TFTP server

Auto Install is successful when a configuration file is downloaded to the switch from a TFTP server.

6.2.1.1. Obtaining IP address Information

DHCP is enabled by default on the service port. If an IP address has not been assigned, the switch issues requests for an IP address assignment.

A network DHCP server returns the following information:

- IP address and subnet mask to be assigned to the interface
- IP address of a default gateway, if needed for IP communication

6.2.1.2. Obtaining Other Dynamic Information

The following information is also processed and may be returned by a BOOTP or DHCP server:

- Name of configuration file (the *file* field in the DHCP header or option 67) to be downloaded from the TFTP server.
- Identification of the TFTP server providing the file. The TFTP server can be identified by name or by IP address as follows:
  - Host name: DHCP option 66 or the *sname* field in the DHCP header
  - IP address: DHCP option 150 or the *siaddr* field in the DHCP header

When a DHCP OFFER identifies the TFTP server more than once, the DHCP client selects one of the options in the following order: *sname*, option 66, option 150, *siaddr*. If the TFTP server is identified by host name, a DNS server is required to translate the name to an IP address.

6.2.1.3. Obtaining the Configuration File

If the DHCP OFFER identifies a configuration file, either as option 67 or in the *file* field of the DHCP header, the switch attempts to download the configuration file.

Note: The configuration file is required to have a file type of *.cfg.

The TFTP client makes three unicast requests. If the unicast attempts fail, or if the DHCP OFFER did not specify a TFTP server address, the auto install process will fail.

6.2.2. Monitoring and Completing the DHCP Auto Install Process

When the switch boots and triggers an Auto Install, a message is written to the buffered log. After the process completes, the Auto Install process writes a log message. You can use the `show logging buffered` command to view information about the process.

Additionally, while the Auto Install is running, you can issue the `show autoinstall` command to view information about the current Auto Install state.

When Auto Install has successfully completed, you can execute a `show running-config` command to validate the contents of configuration.

6.2.2.1. Saving a Configuration

The Auto Install feature includes an AutoSave feature that allows the downloaded configuration to be automatically saved; AutoSave is enabled by default. This makes the configuration available for the next reboot. In the CLI, this is performed by issuing a `copy running-config startup-config` command and should be done after validating the contents of saved configuration.

6.2.2.2. Stopping and Restarting the Auto Install Process

You can terminate the Auto Install process at any time before the configuration file is downloaded. This is useful when the switch is disconnected from the network. Termination of the Auto Install process ends further periodic requests for a host-specific file.

6.2.2.3. Managing Downloaded Configuration Files

The configuration files downloaded to the switch by Auto Install are stored in the nonvolatile memory as `.scr` files. The files may be managed (viewed or deleted) along with files downloaded by the configuration scripting utility. If the Auto Install persistent mode is enabled (`boot-system host autoinstall`) and the switch reboots, the `.scr` configuration file created by the switch in the non-volatile memory is overwritten during the Auto Install process.

To ensure that the downloaded configuration file is used during the next boot cycle, make sure that the Auto Install persistent mode is disabled (`no boot-system host autoinstall`) and save the configuration (`copy running-config startup-config`).

6.2.3. DHCP Auto Install Dependencies

The Auto Install process from TFTP servers depends upon the following network services:

- A DHCP server must be configured on the network with appropriate services.
- A configuration file (either from bootfile (or) option 67 option) for the switch must be available from a TFTP server.
- The switch must be connected to the network and have a Layer 3 interface that is in an UP state.
- A DNS server must contain an IP address to host name mapping for the TFTP server if the DHCP server response identifies the TFTP server by name.
- If a default gateway is needed to forward TFTP requests, an IP helper address for TFTP needs to be configured on the default gateway.

6.2.4. Default Auto Install Values

The following table describes the Auto Install defaults.

Feature	Default	Description
Retry Count	3	When the DHCP or BOOTP server returns information about the TFTP server and bootfile, the switch makes three unicast TFTP requests for the specified bootfile.
AutoSave	Enabled	If the switch is successfully auto-configured, the running configuration is saved to the startup configuration.

Table 6-2: Auto Install Defaults

6.2.5. Enabling DHCP Auto Install

A network administrator is deploying three switches and wants to quickly and automatically install a common configuration file that configures basic settings such as VLAN creation and membership and RADIUS server settings. This example describes the procedures to complete the configuration. The DHCP and TFTP servers in this example are reachable from the service port on the switch.

To use DHCP Auto Install:

1. Create a default config file for the switches named `host.cfg`.

2. Upload the `host.cfg` file to the TFTP server.

3. Configure an address pool on the DHCP server that contains the following information:

a. The IP address (*yiaddr*) and subnet mask (option 1) to be assigned to the interface

b. The IP address of a default gateway (option 3)

c. DNS server address (option 6)

d. Name of config file for each host

e. Identification of the TFTP server by host name (DHCP option 66 or the *sname* field in the DHCP header) or IP address (DHCP option 150 or the *siaddr* field in the DHCP header)

4. Connect the service port on each switch to the management network. This network must have a route to the DHCP server and TFTP server that are used for Auto Install process.

5. Reboot each switch.

```
(Switch) #reload
```

6.3. Configuring System Log Example

In this example, Log server installed 3 party log software and enable log server service on UDP port 514. The log server connects to the service port on Switch A with IP address 172.16.100.240/24.

6.3.1. Example 1 to Add Syslog Host

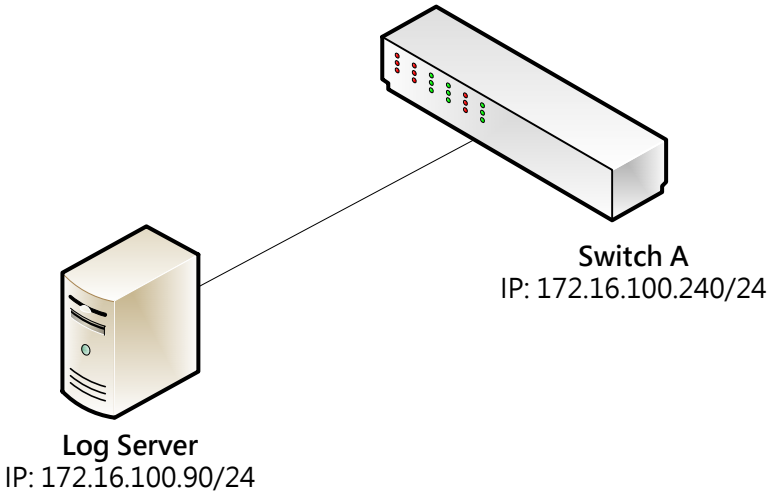


Figure 6-1: System Log Topology

6.3.2. Example 2 to Verify Syslog Host Configuration

To configure Switch A

1.Configure log server ip address 172.16.100.90.

```
(Switch) (Config)#logging host 172.16.100.90 ipv4
```

2.Configure log server received port number to 514.

Note: Default syslog server port is 514

```
(Switch) (Config)#logging host reconfigure 1 port 514
```

3.Change the log severity level to 6

```
(Switch) (Config)#logging host reconfigure 1 severitylevel 6
```

severity level mapping number	
emergency	0
alert	1
critical	2
error	3
warning	4
notice	5
info	6
debug	7

4.Enable syslog feature

```
(Switch) (Config)#logging syslog
```

Result: The syslog server receives log messages from switch A. See the following figure.

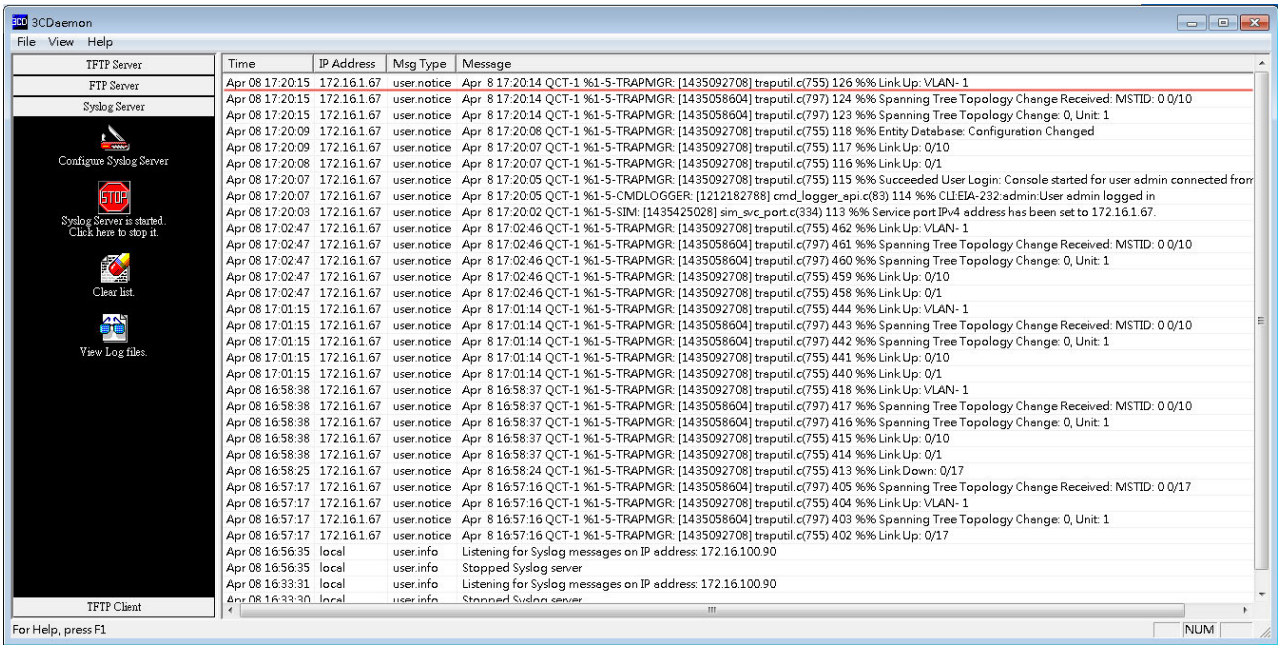


Figure 6-2: Syslog Server Screen

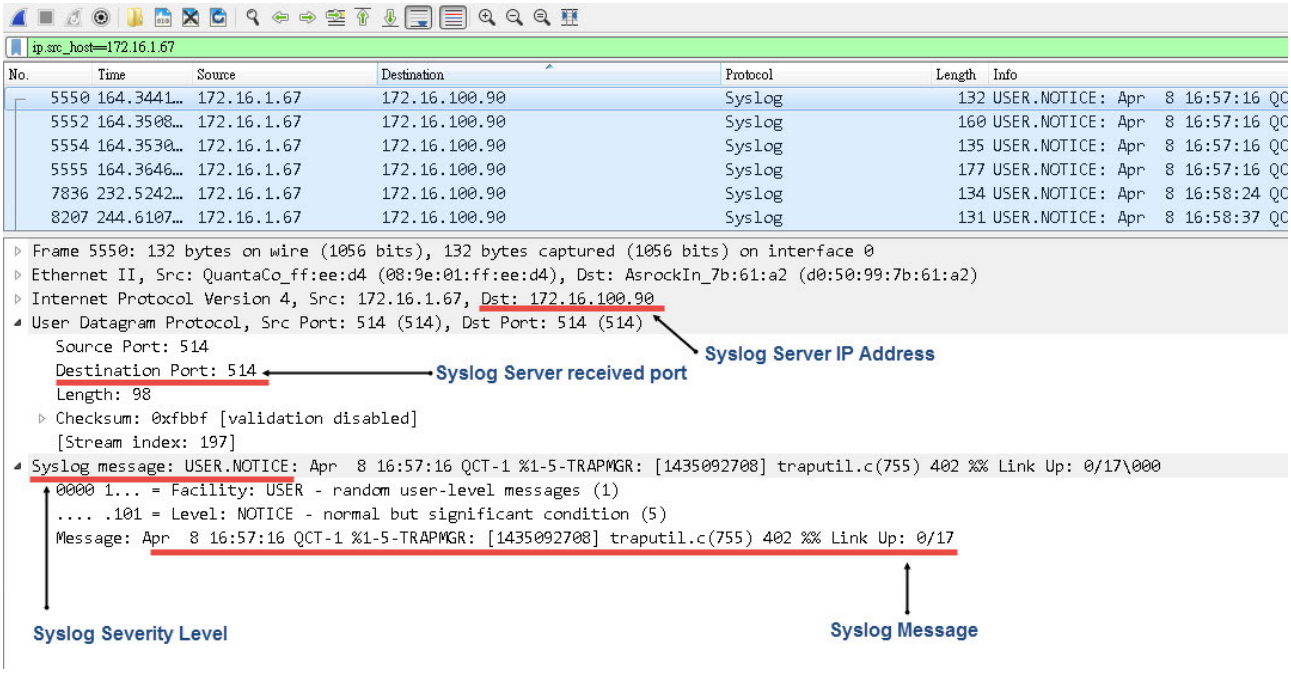


Figure 6-3: Syslog packet capture

Using show logging command to verify the logging configuration

```
(Switch) (Config)#show logging

Logging Client Local Port           : 514

Logging Client Source Interface     : (not configured)

CLI Command Logging                : disabled

Console Logging                     : enabled

Console Logging Severity Filter     : error

Buffered Logging                    : enabled

Buffered Logging Severity Filter    : info

Persistent Logging                  : disabled

Persistent Logging Severity Filter  : alert


Syslog Logging                      : enabled

Syslog Logging Facility             : user


Terminal Monitor                   : disabled

Terminal Logging Severity Filter    : warning


Log Messages Received               : 246

Log Messages Dropped                : 0
```

```
Log Messages Relayed                : 0
```

Using show logging hosts command to verify the logging hosts configuration

```
(Switch-1) #show logging hosts

Index   IP Address/Hostname      Type     Severity    Port    Status
-----   --------------------    -------  ----------  ------  ----------
1      172.16.100.90            ipv4      notice     514    Active
```

Example3: Remove Syslog host

Using show logging hosts to check the syslog server index

```
(Switch-1) #show logging hosts

Index  IP Address/Hostname      Type     Severity    Port    Status
-----   -------------------    -------  ----------  ------   -------------
1      172.16.100.90            ipv4      notice     514    Active
2      172.16.100.230           ipv4      notice     514    Active
```

Using remove command to remove syslog server 2

```
(Switch) (Config)#logging host remove 2
```

Result: the syslog server 2 is removed from the syslog server list.

6.4. Configuring CLI Scheduler (Kron)

Configuring CLI scheduler Policy Lists and Occurrences

An occurrence for CLI scheduler is defined as a scheduled event. Policy lists are configured to run after a specified interval of time, at a specified calendar date and time, or upon system startup. Policy lists can be run once, as a one-time event, or as recurring events over time.

CLI scheduler occurrences can be scheduled before the associated policy list has been configured, but a warning will advise you to configure the policy list before it is scheduled to run.

6.4.1. CLI Scheduler Policy Lists

Policy lists consist of one or more lines of fully-qualified EXEC CLI commands. All commands in a policy list are executed when the policy list is run by CLI scheduler using the **kron occurrence** command. Use separate policy lists for CLI commands that are run at different times. No editor function is available, and the policy list is run in the order in which it was configured. To delete an entry, use the **no** form of the **cli** command followed by the appropriate EXEC command. If an existing policy list name is used, new entries are added to the end of the policy list. To view entries in a policy list, use the **show running-config** command. If a policy list is scheduled to run only once, it will not be displayed by the **show running-config** command after it has run.

Policy lists can be configured after the policy list has been scheduled, but each policy list must be configured before it is scheduled to run.

6.4.2. CLI Scheduler Occurrences

The EXEC CLI to be run by CLI scheduler must be tested on the device to determine if it will run without generating a prompt or allowing execution interruption by keystrokes. Initial testing is important because CLI scheduler will delete the entire policy list if any CLI syntax fails. Removing the policy list ensures that any CLI dependencies will not generate more errors.

6.4.3. Configuration Example

The following example guide you how to using CLI Scheduler to backup configuration every week.

1. Create a CLI scheduler policy called BACKUP.

```
(Switch) (Config)#kron policy-list BACKUP
```

2. Configure backup command into BACKUP.

```
(Switch) (Kron-policy)#cli show running-config | redirect tftp://192.168.1.100/switch.cfg
(Switch) (Kron-policy)#exit
```

3. Create occurrence z1 with policy “BACKUP” at Monday 11:13 for a week recurring.

```
(Switch) (Config)#kron occurrence Z1 at 11:13 mon recurring
(Switch) (Kron-occurrence)#policy-list BACKUP
(Switch) (Kron-occurrence)#exit
```

7. Configuring Routing

7.1. Basic Routing and Features

The switch runs on multilayer switches that support static and dynamic routing. The following table describes some of the general routing features that you can configure on the switch. The table does not list supported routing protocols.

Feature	*Description*
ICMP message control	You can configure the type of ICMP messages that the switch responds to as well as the rate limit and burst size.
Default gateway	The switch supports a single default gateway. A manually configured default gateway is more preferable than a default gateway learned from a DHCP server.
ARP table	The switch maintains an ARP table that maps an IP address to a MAC address. You can create static ARP entries in the table and manage various ARP table settings such as the aging time of dynamically-learned entries.
Routing table entries	You can configure the following route types in the routing table: • Default: The default route is the route the switch will use to send a packet if the routing table does not contain a longer matching prefix for the packet's destination. • Static: A static route is a route that you manually add to the routing table. • Static Reject: Packets that match a reject route are discarded instead of forwarded. The router may send an ICMP Destination Unreachable message.
Route preferences	The common routing table collects static, local, and dynamic (routing protocol) routes. When there is more than one route to the same destination prefix, the routing table selects the route with the best (lowest) route preference.

Table 7-1: IP Routing Features

7.1.1. VLAN Routing

VLANs divide a single physical network (broadcast domain) into separate logical networks. To forward traffic across VLAN boundaries, a layer 3 device, such as router, is required. The switch can act as a layer 3 device when you configure VLAN routing interfaces. VLAN routing interfaces make it possible to transmit traffic between VLANs while still containing broadcast traffic within VLAN boundaries. The configuration of VLAN routing interfaces makes inter-VLAN routing possible.

For each VLAN routing interface you can assign a static IP address, or you can allow a network DHCP server to assign a dynamic IP address.

When a port is enabled for bridging (L2 switching) rather than routing, which is the default, all normal bridge processing is performed for an inbound packet, which is then associated with a VLAN. It's MAC Destination Address (MAC DA) and VLAN ID are used to search the MAC address table. If routing is enabled for the VLAN, and the MAC DA of an inbound unicast packet is that of the internal router interface, the packet is routed. An inbound multicast packet is forwarded to all ports in the VLAN, plus the internal bridge-router interface, if it was received on a routed VLAN.

Since a port can be configured to belong to more than one VLAN, VLAN routing might be enabled for all of the VLANs on the port or for only some of the VLANs on the port. VLAN Routing can be used to allow more than

one physical port to reside on the same subnet. It could also be used when a VLAN spans multiple physical networks, or when additional segmentation or security is required.

7.1.1.1. When to Configure VLAN Routing

VLAN routing is required when the switch is used as a layer 3 device. VLAN routing must be configured to allow the switch to forward IP traffic between subnets and allow hosts in different networks to communicate.

In the following figure, the M4500 series switch is configured as an L3 device and performs the routing functions for hosts connected to the L2 switches. For Host A to communicate with Host B, no routing is necessary. These hosts are in the same VLAN. However, for Host A in VLAN 10 to communicate with Host C in VLAN 20, the switch must perform inter-VLAN routing.

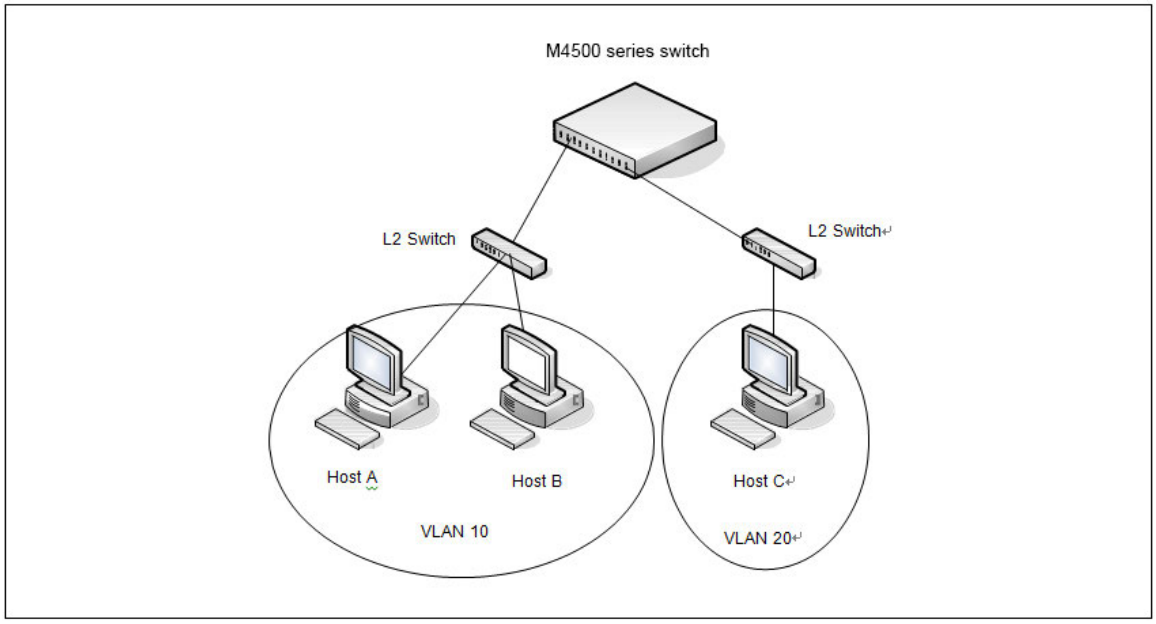


Figure 7-1: Inter-VLAN Routing

7.1.2. IP Routing Configuration Example

In this example, the switches are L3 switches with VLAN routing interfaces. VLAN routing is configured on Switch A and Switch B. This allows the host in VLAN 10 to communicate with the server in VLAN 30. A static route to the VLAN 30 subnet is configured on Switch A. Additionally, a default route is configured on Switch A so that all traffic with an unknown destination is sent to the backbone router through port 24, which is a member of VLAN 50. A default route is configured on Switch B to use Switch A as the default gateway. The hosts use the IP address of the VLAN routing interface as their default gateway.

This example assumes that all L2 VLAN information, such as VLAN creation and port membership, has been configured.

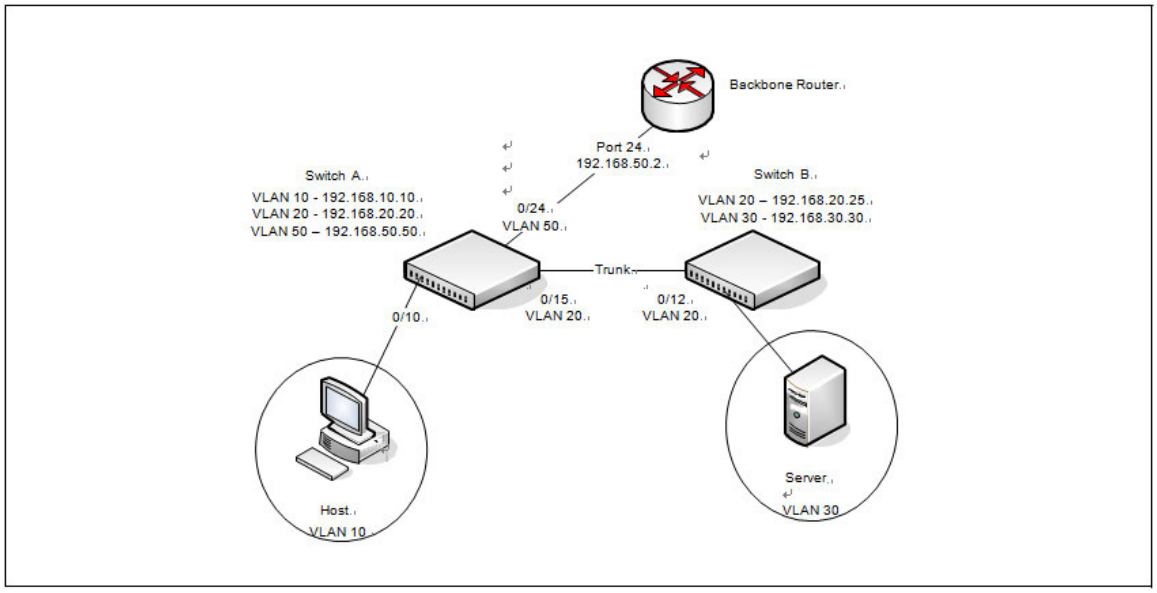


Figure 7-2: IP Routing Example Topology

7.1.2.1. Configuring Switch A

To configure Switch A.

1. Create the VLANs.

```
(Switch) #configure
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 10,20,30,50
```

2. Configure the VLANs for routing.

```
(Switch) (Config)#interface vlan 10
Interface vlan 10 created for VLAN ID 10
(Switch) (if-vlan10)#interface vlan 20
Interface vlan 20 created for VLAN ID 20
(Switch) (if-vlan20)#interface vlan 30
Interface vlan 30 created for VLAN ID 30
(Switch) (if-vlan30)#interface vlan 50
Interface vlan 50 created for VLAN ID 50
(Switch) (if-vlan50)#exit
```

3. Enable routing on the switch.

```
(Switch) #configure
(Switch) (Config)#ip routing
```

4. Assign an IP address to VLAN 10. This command also enables IP routing on the VLAN.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip address 192.168.10.10 255.255.255.0
(Switch) (if-vlan10)#exit
```

5. Assign an IP address to VLAN 20.

```
(Switch) (Config)#interface vlan 20
(Switch) (if-vlan20)#ip address 192.168.20.20 255.255.255.0
(Switch) (if-vlan20)#exit
```

6. Assign an IP address to VLAN 50.

```
(Switch) (Config)#interface vlan 50
(Switch) (if-vlan50)#ip address 192.168.50.50 255.255.255.0
(Switch) (if-vlan50)#exit
```

7. Configure a static route to the network that VLAN 30 is in, using the IP address of the VLAN 20 interface on Switch B as the next hop address.

```
(Switch) (Config)#ip route 192.168.30.0 255.255.255.0 192.168.20.25
```

8. Configure the backbone router interface as the default gateway.

```
(Switch) (Config)#ip route default 192.168.50.2
```

7.1.2.2. Configuring Switch B

To configure Switch B:

1. Create the VLANs.

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 20,30
```

2. Configure the VLANs for routing.

```
(Switch) (Config)#interface vlan 20
Interface vlan 20 created for VLAN ID 20
(Switch) (if-vlan20)#interface vlan 30
Interface vlan 30 created for VLAN ID 30
(Switch) (if-vlan30)#exit
```

3. Enable routing on the switch.

```
(Switch) #configure
(Switch) (Config)#ip routing
```

4. Assign an IP address to VLAN 20. This command also enables IP routing on the VLAN.

```
(Switch) (Config)#interface vlan 20
(Switch) (if-vlan20)#ip address 192.168.20.25 255.255.255.0
(Switch) (if-vlan20)#exit
```

5. Assign an IP address to VLAN 30. This command also enables IP routing on the VLAN.

```
(Switch) (Config)#interface vlan 30
(Switch) (if-vlan30)#ip address 192.168.30.30 255.255.255.0
(Switch) (if-vlan30)#exit
```

6. Configure the VLAN 20 routing interface on Switch A as the default gateway so that any traffic with an unknown destination is sent to Switch A for forwarding.

```
(Switch) (Config)#ip route default 192.168.20.20
```

7.1.3. IP Unnumbered Configuration Example

This IP unnumbered configuration example shows how the same IP is used on two different unnumbered interfaces on Router 1 so it can communicate with Router 2 and Router 3.

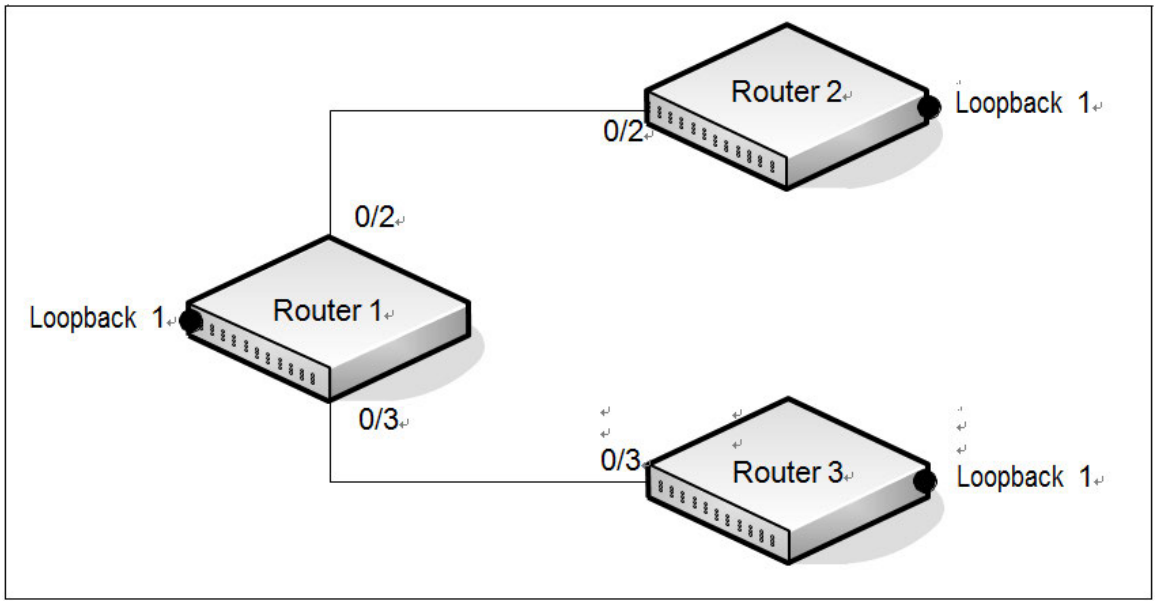


Figure 7-3: IP Unnumbered Configuration Example

To configure Router 1:

1. Enable routing on the switch.

```
(Switch) #configure
(Switch) (Config)#ip routing
```

2. Configure the loopback interface.

```
(Switch) (Config)#interface loopback 1
(Switch) (Interface loopback 1)#ip address 1.0.0.1 /24
(Switch) (Interface loopback 1)#exit
```

3. Configure port 0/2.

```
(Switch) (Config)#interface 0/2
(Switch) (Interface 0/2)#routing
(Switch) (Interface 0/2)#ip unnumbered loopback 1
(Switch) (Interface 0/2)#exit
```

4. Configure port 0/3.

```
(Switch) (Interface 0/3)#routing
(Switch) (Interface 0/3)#ip unnumbered loopback 1
(Switch) (Interface 0/3)#exit
(Switch) (Config)#
```

To configure Router 2:

1. Enable routing on the switch.

```
(Switch) #configure
(Switch) (Config)#ip routing
```

2. Configure the loopback interface.

```
(Switch) (Config)#interface loopback 1
(Switch) (Interface loopback 1)#ip address 2.0.0.2 /24
(Switch) (Interface loopback 1)#exit
```

3. Configure port 0/2.

```
(Switch) (Config)#interface 0/2
(Switch) (Interface 0/2)#routing
(Switch) (Interface 0/2)#ip unnumbered loopback 1
(Switch) (Interface 0/2)#exit
```

4. Configure port 0/3.

```
(Switch) (Interface 0/3)#routing
(Switch) (Interface 0/3)#ip unnumbered loopback 1
(Switch) (Interface 0/3)#exit
(Switch) (Config)#
```

To configure Router 3:

1. Enable routing on the switch.

```
(Switch) #configure
(Switch) (Config)#ip routing
```

2. Configure the loopback interface.

```
(Switch) (Config)#interface loopback 1
(Switch) (Interface loopback 1)#ip address 3.0.0.3 /24
(Switch) (Interface loopback 1)#exit
```

3. Configure port 0/2.

```
(Switch) (Config)#interface 0/2
(Switch) (Interface 0/2)#routing
(Switch) (Interface 0/2)#ip unnumbered loopback 1
(Switch) (Interface 0/2)#exit
```

4. Configure port 0/3.

```
(Switch) (Interface 0/3)#routing
(Switch) (Interface 0/3)#ip unnumbered loopback 1
(Switch) (Interface 0/3)#exit
(Switch) (Config)#
```

When you have completed the configuration instructions above, try to ping 2.0.0.2 and 3.0.0.3 from router 1.

7.2. OSPF

OSPF is an Interior Gateway Protocol (IGP) that performs dynamic routing within a network. The top level of the hierarchy of an OSPF network is known as an OSPF domain. The domain can be divided into areas. Routers within an area must share detailed information on the topology of their area, but require less detailed information about the topology of other areas. Segregating a network into areas enables limiting the amount of route information communicated throughout the network.

Areas are identified by a numeric ID in IP address format n.n.n.n (note, however, that these are not used as actual IP addresses). For simplicity, the area can be configured and referred to in normal integer notation. For example, Area 20 is identified as 0.0.0.20 and Area 256 as 0.0.1.0. The area identified as 0.0.0.0 is referred to as Area 0 and is considered the OSPF backbone. All other OSPF areas in the network must connect to Area 0 directly or through a virtual link. The backbone area is responsible for distributing routing information between non-backbone areas.

A virtual link can be used to connect an area to Area 0 when a direct link is not possible. A virtual link traverses an area between the remote area and Area 0.

A stub area is an area that does not accept external LSAs (LSAs generated by redistributing routes) that were learned from a protocol other than OSPF or were statically configured. These routes typically send traffic outside the AS. Therefore, routes from a stub area to locations outside the AS use the default gateway. A virtual link cannot be configured across a stub area. A Not So Stubby Area can import limited external routes only from a connected ASBR.

7.2.1. Configuring an OSPF Border Router and Setting Interface Costs

This example shows how to configure the switch as an OSPF border router. The commands in this example configure the areas and interfaces on Border Router A shown in the following figure.

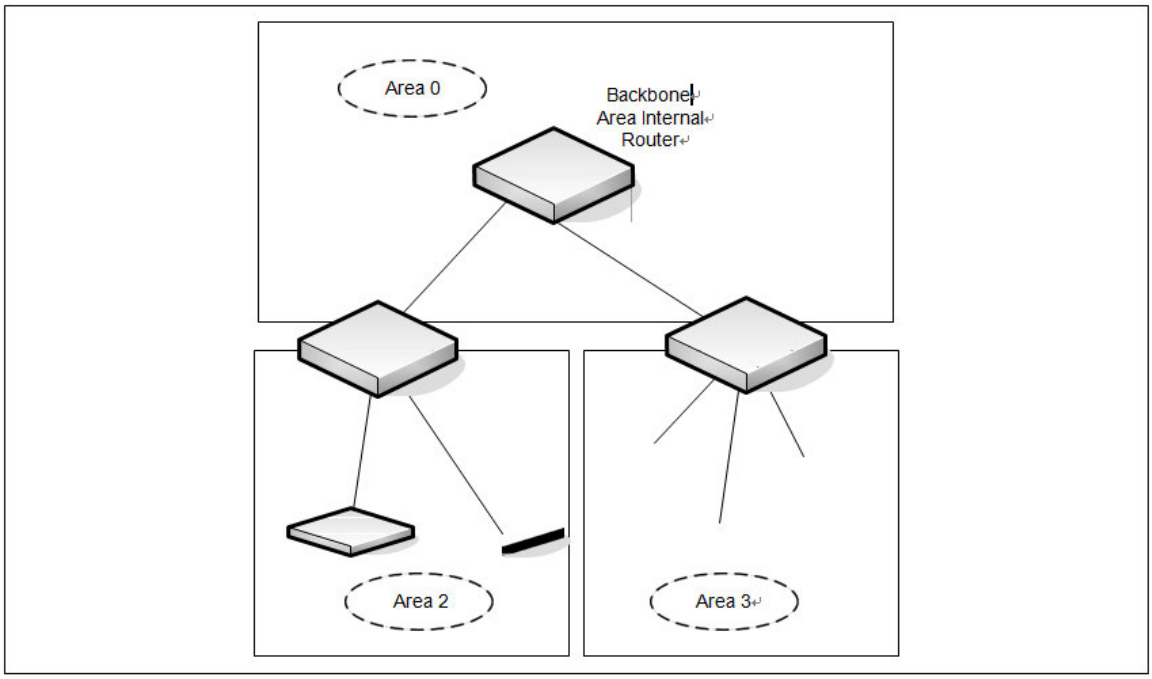


Figure 7-4: OSPF Area Border Router

To Configure Border Router A:

1. Enable routing on the switch.

```
(Switch) #configure
(Switch) (Config)#ip routing
```

2. Create VLANS 70, 80, and 90.

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 70,80,90
```

3. Configure the VLANs for routing and assign the interface port numbers.

```
(Switch) (Config)#interface vlan 70
Interface vlan 70 created for VLAN ID 70
(Switch) (if-vlan70)#interface vlan 80
Interface vlan 80 created for VLAN ID 80
(Switch) (if-vlan80)#interface vlan 90
Interface vlan 90 created for VLAN ID 90
(Switch) (if-vlan50)#exit
```

4. Enable routing on the switch.

```
(Switch) #configure
(Switch) (Config)#ip routing
```

5. Assign IP addresses for VLANs 70, 80 and 90.

```
(Switch) (Config)#interface vlan 70
(Switch) (if-vlan70)#ip address 192.150.2.2 255.255.255.0
(Switch) (if-vlan70)#exit

(Switch) (Config)#interface vlan 80
(Switch) (if-vlan80)#ip address 192.150.3.1 255.255.255.0
(Switch) (if-vlan80)#exit

(Switch) (Config)#interface vlan 90
(Switch) (if-vlan90)#ip address 192.150.4.1 255.255.255.0
(Switch) (if-vlan90)#exit
```

6. Enable OSPF on the switch and specify a router ID.

```
(Switch) (Config)#router ospf
(Switch) (Config-router)#router-id 192.150.9.9
(Switch) (Config-router)#exit
```

7. Configure the OSPF area ID and cost for each interface.

Note: OSPF is globally enabled by default. To make it operational on the router, you configure OSPF for particular interfaces and identify which area the interface is associated with.

```
(Switch) (Config)#interface vlan 70
(Switch) (if-vlan70)#ip ospf area 0.0.0.0
(Switch) (if-vlan70)#ip ospf cost 32
(Switch) (if-vlan70)#exit
```

```
(Switch) (Config)#interface vlan 80
(Switch) (if-vlan80)#ip ospf area 0.0.0.2
(Switch) (if-vlan80)#ip ospf cost 64
(Switch) (if-vlan80)#exit

(Switch) (Config)#interface vlan 90
(Switch) (if-vlan90)#ip ospf area 0.0.0.2
(Switch) (if-vlan90)#ip ospf cost 64
(Switch) (if-vlan90)#exit
```

7.3. VRRP

The Virtual Router Redundancy (VRRP) protocol is designed to handle default router (L3 switch) failures by providing a scheme to dynamically elect a backup router. VRRP can help minimize black hole periods due to the failure of the default gateway router during which all traffic directed towards it is lost until the failure is detected.

7.3.1. VRRP Operation in the Network

VRRP eliminates the single point of failure associated with static default routes by enabling a backup router to take over from a master router without affecting the end stations using the route. The end stations will use a virtual IP address that will be recognized by the backup router if the master router fails. Participating routers use an election protocol to determine which router is the master router at any given time. A given port may appear as more than one virtual router to the network, also, more than one port on a switch may be configured as a virtual router. Either a physical port or a routed VLAN may participate.

With VRRP, a virtual router is associated with one or more IP addresses that serve as default gateways. In the event that the VRRP router controlling these IP addresses (formally known as the master) fails, the group of IP addresses and the default forwarding role is taken over by a Backup VRRP Router.

7.3.1.1. VRRP Router Priority

The VRRP router priority is a value from 1–255 that determines which router is the master. The greater the number, the higher the priority. If the virtual IP address is the IP address of a VLAN routing interface on one of the routers in the VRRP group, the router with IP address that is the same as the virtual IP address is the interface owner and automatically has a priority of 255. By default, this router is the VRRP master in the group.

If no router in the group owns the VRRP virtual IP address, the router with the highest configured priority is the VRRP master. If multiple routers have the same priority, the router with the highest IP address becomes the VRRP master.

If the VRRP master fails, other members of the VRRP group will elect a master based on the configured router priority values. For example, router A is the interface owner and master, and it has a priority of 255. Router B is

configured with a priority of 200, and Router C is configured with a priority of 190. If Router A fails, Router B assumes the role of VRRP master because it has a higher priority.

7.3.1.2. VRRP Preemption

If preempt mode is enabled and a router with a higher priority joins the VRRP group, it takes over the VRRP master role if the current VRRP master is not the owner of the virtual IP address. The preemption delay controls how long to wait to determine whether a higher priority Backup router preempts a lower priority Master. In certain cases, for example, during periods of network congestion, a backup router might fail to receive advertisements from the master. This could cause members in the VRRP group to change their states frequently, i.e. flap. The problem can be resolved by setting the VRRP preemption delay timer to a non-zero value.

7.3.1.3. VRRP Accept Mode

The accept mode allows the switch to respond to pings (ICMP Echo Requests) sent to the VRRP virtual IP address. The VRRP specification (RFC 3768) indicates that a router may accept IP packets sent to the virtual router IP address only if the router is the address owner. In practice, this restriction makes it more difficult to troubleshoot network connectivity problems. When a host cannot communicate, it is common to ping the host's default gateway to determine whether the problem is in the first hop of the path to the destination. When the default gateway is a virtual router that does not respond to pings, this troubleshooting technique is unavailable. On the switch, you can enable Accept Mode to allow the system to respond to pings that are sent to the virtual IP address.

This capability adds support for responding to pings, but does not allow the VRRP Master to accept other types of packets. The VRRP Master responds to both fragmented and un-fragmented ICMP Echo Request packets. The VRRP Master responds to Echo Requests sent to the virtual router's primary address or any of its secondary addresses.

Members of the virtual router who are in backup state discard ping packets destined to VRRP addresses, just as they discard any Ethernet frame sent to a VRRP MAC address.

When the VRRP master responds with an Echo Reply, the source IPv4 address is the VRRP address and source MAC address is the virtual router's MAC address.

7.3.1.4. VRRP Route and Interface Tracking

The VRRP Route/Interface Tracking feature extends VRRP capability to allow tracking of specific routes and interface IP states within the router that can alter the priority level of a virtual router for a VRRP group.

VRRP interface tracking monitors a specific interface IP state within the router. Depending on the state of the tracked interface, the feature can alter the VRRP priority level of a virtual router for a VRRP group.

Note: An exception to the priority level change is that if the VRRP group is the IP address owner, its priority is fixed at 255 and cannot be reduced through the tracking process.

With standard VRRP, the backup router takes over only if the router goes down. With VRRP interface tracking, if a tracked interface goes down on the VRRP master, the priority decrement value is subtracted from the router

priority. If the master router priority becomes less than the priority on the backup router, the backup router takes over. If the tracked interface becomes up, the value of the priority decrement is added to the current router priority. If the resulting priority is more than the backup router priority, the original VRRP master resumes control.

VRRP route tracking monitors the reachability of an IP route. A tracked route is considered up when a routing table entry exists for the route and the route is accessible. When the tracked route is removed from the routing table, the priority of the VRRP router will be reduced by the priority decrement value. When the tracked route is added to the routing table, the priority will be incremented by the same.

7.3.2. VRRP Configuration Example

This section contains the following VRRP examples:

- VRRP with Load Sharing
- VRRP with Route and Interface Tracking

7.3.2.1. VRRP with Load Sharing

In the following figure, two L3 switches are performing the routing for network clients. Router A is the default gateway for some clients, and Router B is the default gateway for other clients.

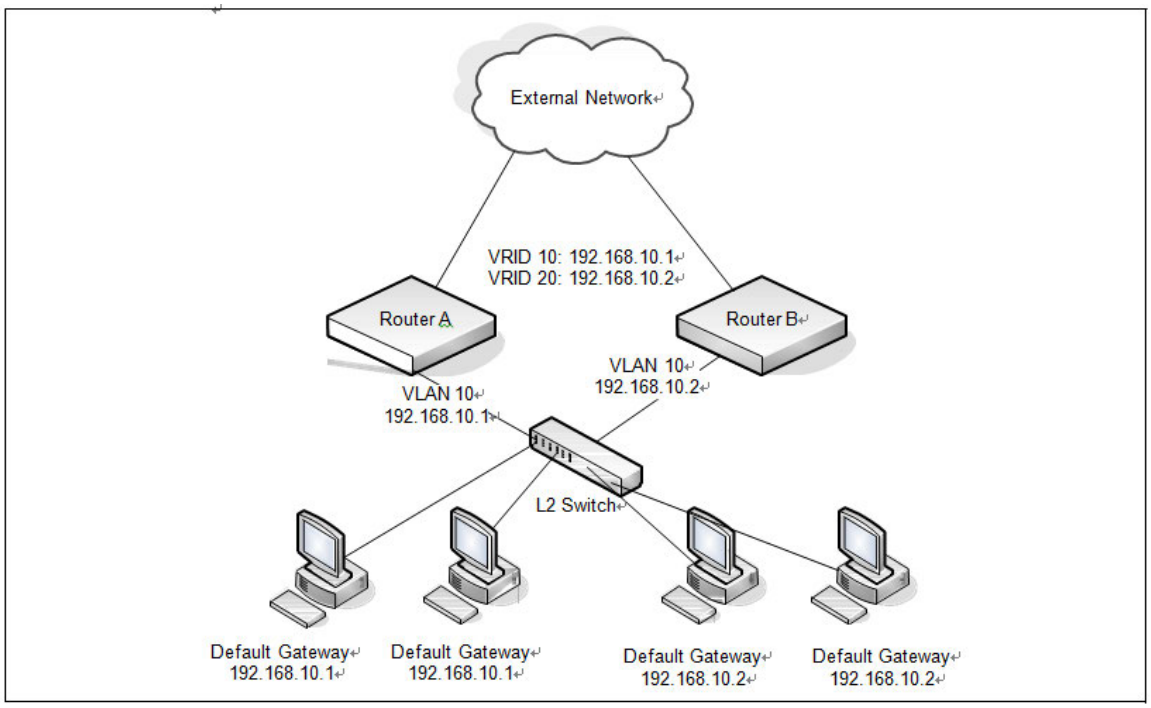


Figure 7-5: VRRP with Load Sharing Network Diagram

This example configures two VRRP groups on each router. Router A is the VRRP master for the VRRP group with VRID 10 and the backup for VRID 20. Router B is the VRRP master for VRID 20 and the backup for VRID 10.

If Router A fails, Router B will become the master of VRID 10 and will use the virtual IP address 192.168.10.1. Traffic from the clients configured to use Router A as the default gateway will be handled by Router B.

To configure Router A:

1. Create and configure the VLAN routing interface to use as the default gateway for network clients. This example assumes all other routing interfaces, such as the interface to the external network, have been configured.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip address 192.168.10.1 255.255.255.0
(Switch) (if-vlan10)#exit
```

2. Enable routing for the switch.

```
(Switch) (Config)#ip routing
```

3. Enable VRRP for the switch.

```
(Switch) (Config)#ip vrrp
```

4. Assign a virtual router ID to the VLAN routing interface for the first VRRP group.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip vrrp 10
```

5. Specify the IP address that the virtual router function will use. The router is the virtual IP address owner (the routing interface has the same IP address as the virtual IP address for the VRRP group), so the priority value is 255.

```
(Switch) (if-vlan10)#ip vrrp 10 ip 192.168.10.1
```

6. Assign a virtual router ID to the VLAN routing interface for the second VRRP group.

```
(Switch) (if-vlan10)#ip vrrp 20
```

7. Specify the IP address that the virtual router function will use.

```
(Switch) (if-vlan10)#ip vrrp 20 ip 192.168.10.2
```

8. Enable the VRRP groups on the interface.

```
(Switch) (if-vlan10)#ip vrrp 10 mode
(Switch) (if-vlan10)#ip vrrp 20 mode
(Switch) (if-vlan10)#exit
(Switch) (Config)#exit
```

The only difference between the Router A and Router B configurations is the IP address assigned to VLAN 10. On Router B, the IP address of VLAN 10 is 192.168.10.2. Because this is also the virtual IP address of VRID 20, Router B is the interface owner and VRRP master of VRRP group 20.

To configure Router B:

1. Enable routing for the switch.

```
(Switch) #config
(Switch) (Config)#ip routing
(Switch) (Config)#exit
```

2. Create and configure the VLAN routing interface to use as the default gateway for network clients. This example assumes all other routing interfaces, such as the interface to the external network, have been configured.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip address 192.168.10.2 255.255.255.0
(Switch) (if-vlan10)#exit
```

3. Enable VRRP for the switch.

```
(Switch) (Config)#ip vrrp
```

4. Assign a virtual router ID to the VLAN routing interface for the first VRRP group.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip vrrp 10
```

5. Specify the IP address that the virtual router function will use.

```
(Switch) (if-vlan10)#ip vrrp 10 ip 192.168.10.1
```

6. Assign a virtual router ID to the VLAN routing interface for the second VRRP group.

```
(Switch) (if-vlan10)#ip vrrp 20
```

7. Specify the IP address that the virtual router function will use.

The router is the virtual IP address owner of this address, so the priority value is 255 by default.

```
(Switch) (if-vlan10)#ip vrrp 20 ip 192.168.10.2
```

8. Enable the VRRP groups on the interface.

```
(Switch) (if-vlan10)#ip vrrp 10 mode
(Switch) (if-vlan10)#ip vrrp 20 mode
(Switch) (if-vlan10)#exit
(Switch) (Config)#exit
```

7.3.2.2. VRRP with Route and Interface Tracking

In the following figure, the VRRP priorities are configured so that Router A is the VRRP master, and Router B is the VRRP backup. Router A forwards IP traffic from clients to the external network through the VLAN 25 routing interface. The clients are configured to use the virtual IP address 192.168.10.15 as the default gateway.

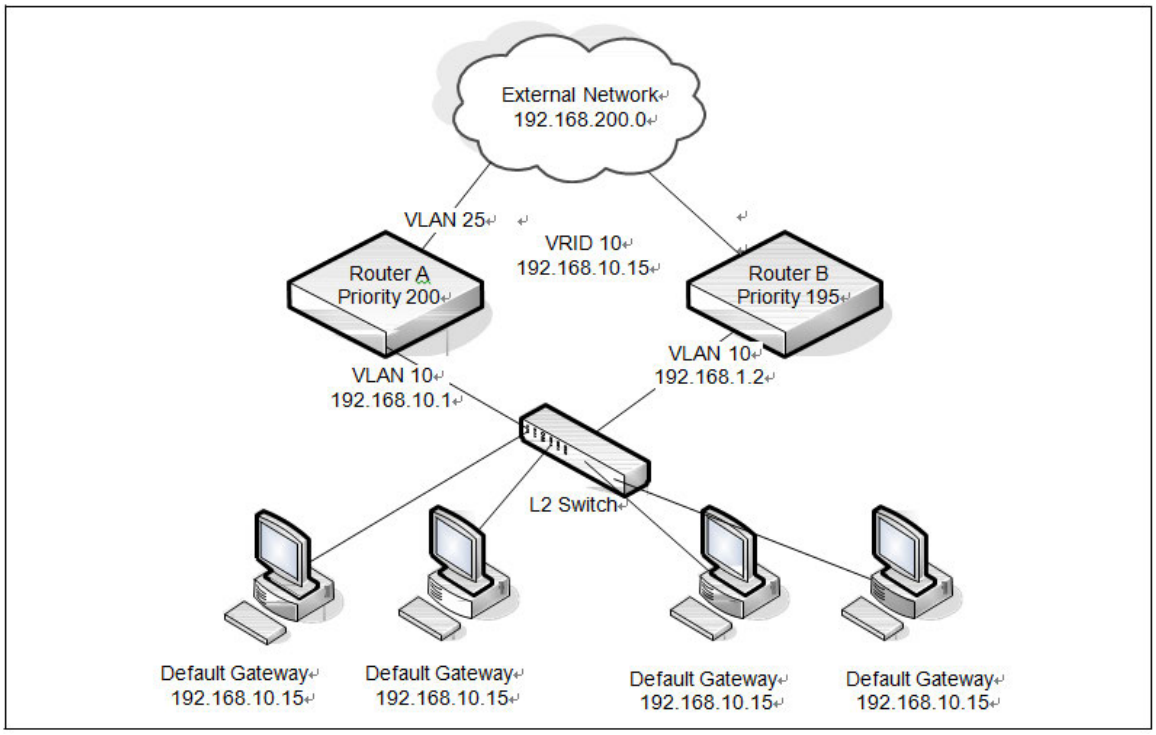


Figure 7-6: VRRP with Tracking Network Diagram

Without VRRP interface or route tracking, if something happened to VLAN 25 or the route to the external network, as long as Router A remains up, it will continue to be the VRRP master even though traffic from the clients does not have a path to the external network. However, if the interface and/or route tracking features are configured, Router A can decrease its priority value when the problems occur so that Router B becomes the master.

To configure Router A:

1. Enable routing for the switch.

```
(Switch) #config
(Switch) (Config)#ip routing
(Switch) (Config)#exit
```

2. Configure the VLAN routing interface to use as the default gateway for network clients. This example assumes all other routing interfaces, such as the interface to the external network, have been configured.

```
(Switch) #config
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip address 192.168.10.1 255.255.255.0
(Switch) (if-vlan10)#exit
```

3. Enable VRRP for the switch.

```
(Switch) (Config)#ip vrrp
```

4. Assign a virtual router ID to the VLAN routing interface for the VRRP group.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip vrrp 10
```

5. Specify the IP address that the virtual router function will use.

```
(Switch) (if-vlan10)#ip vrrp 10 ip 192.168.10.15
```

6. Configure the router priority.

```
(Switch) (if-vlan10)#ip vrrp 10 priority 200
```

7. Enable preempt mode so that the router can regain its position as VRRP master if its priority is greater than the priority of the backup router.

```
(Switch) (if-vlan10)#ip vrrp 10 preempt
```

8. Enable the VRRP groups on the interface.

```
(Switch) (if-vlan10)#ip vrrp 10 mode
(Switch) (if-vlan10)#exit
```

9. Track the routing interface VLAN 25 on VRID 10 so that if it goes down, the priority of VRID 10 on Router A is decreased by 10, which is the default decrement priority value.

```
(Switch) (if-vlan10)#ip vrrp 10 track interface vlan 25
```

10. Track the route to the 192.168.200.0 network. If it becomes unavailable, the priority of VRID 10 on Router A is decreased by 10, which is the default decrement priority value.

```
(Switch) (if-vlan10)#ip vrrp 10 track ip route 192.168.200.0/24
(Switch) (if-vlan10)#exit
```

Router B is the backup router for VRID 10. The configured priority is 195. If the VLAN 25 routing interface or route to the external network on Router A goes down, the priority of Router A will become 190 (or 180, if both the interface and router are down). Because the configured priority of Router B is greater than the actual priority of Router A, Router B will become the master for VRID 10. When VLAN 25 and the route to the external network are back up, the priority of Router A returns to 200, and it resumes its role as VRRP master.

To configure Router B:

1. Enable routing for the switch.

```
(Switch) #config
(Switch) (Config)#ip routing
(Switch) (Config)#exit
```

2. Create and configure the VLAN routing interface to use as the default gateway for network clients. This example assumes all other routing interfaces, such as the interface to the external network, have been configured.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip address 192.168.10.2 255.255.255.0
(Switch) (if-vlan10)#exit
```

3. Enable VRRP for the switch.

```
(Switch) (Config)#ip vrrp
```

4. Assign a virtual router ID to the VLAN routing interface for the VRRP group.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip vrrp 10
```

5. Specify the IP address that the virtual router function will use.

```
(Switch) (if-vlan10)#ip vrrp 10 ip 192.168.10.15
```

6. Configure the router priority.

```
(Switch) (if-vlan10)#ip vrrp 10 priority 195
```

7. Enable preempt mode so that the router can regain its position as VRRP master if its priority is greater than the priority of the backup router.

```
(Switch) (if-vlan10)#ip vrrp 10 preempt
```

8. Enable the VRRP groups on the interface.

```
(Switch) (if-vlan10)#ip vrrp 10 mode
(Switch) (if-vlan10)#exit
(Switch) (Config)#exit
```

7.4. IP Helper

The IP Helper feature provides the ability for a router to forward configured UDP broadcast packets to a particular IP address. This allows applications to reach servers on non-local subnets. This is possible even when the application is designed to assume a server is always on a local subnet or when the application uses broadcast packets to reach the server (with the limited broadcast address 255.255.255.255, or a network directed broadcast address).

You can configure relay entries globally and on routing interfaces. Each relay entry maps an ingress interface and destination UDP port number to a single IPv4 address (the helper address). Multiple relay entries may be configured for the same interface and UDP port, in which case the relay agent relays matching packets to each server address. Interface configuration takes priority over global configuration. If the destination UDP port for a packet matches any entry on the ingress interface, the packet is handled according to the interface configuration. If the packet does not match any entry on the ingress interface, the packet is handled according to the global IP helper configuration.

You can configure discard relay entries. Discard entries are used to discard packets received on a specific interface when those packets would otherwise be relayed according to a global relay entry. Discard relay entries may be configured on interfaces, but are not configured globally.

Additionally, you can configure which UDP ports are forwarded. Certain UDP port numbers can be specified by name in the CLI, but you can also configure a relay entry with any UDP port number. You may configure relay entries that do not specify a destination UDP port. The relay agent assumes that these entries match packets with the UDP destination ports listed in the following table (the list of default ports).

Protocol	*UDP Port Number*
IEN-116 Name Service	42
DNS	53
NetBIOS Name Server	137
NetBIOS Datagram Server	138
TACACS Server	49
Time Service	37
DHCP	67
Trivial File Transfer Protocol	69

Table 7-2: Default Ports – UDP Port Numbers Implied by Wildcard

The system limits the number of relay entries to four times the maximum number of routing interfaces (512 relay entries). There is no limit to the number of relay entries on an individual interface, and no limit to the number of servers for a given {interface, UDP port} pair.

Certain configurable DHCP relay options do not apply to relay of other protocols. You may optionally set a maximum hop count or minimum wait time using the `bootpdhcprelay maxhopcount` and `bootpdhcprelay minwaittime` commands.

The relay agent relays DHCP packets in both directions. It relays broadcast packets from the client to one or more DHCP servers, and relays packets to the client that the DHCP server unicasts back to the relay agent. For other protocols, the relay agent only relays broadcast packets from the client to the server. Packets from the server back to the client are assumed to be unicast directly to the client. Because there is no relay in the return direction for protocols other than DHCP, the relay agent retains the source IP address from the original client packet. The relay agent uses a local IP address as the source IP address of relayed DHCP client packets.

When a switch receives a broadcast UDP packet on a routing interface, the relay agent verifies that the interface is configured to relay to the destination UDP port. If so, the relay agent unicasts the packet to the configured server IP addresses. Otherwise, the relay agent verifies that there is a global configuration for the destination UDP port. If so, the relay agent unicasts the packet to the configured server IP addresses. Otherwise the packet is not relayed.

Note: If the packet matches a discard relay entry on the ingress interface, the packet is not forwarded, regardless of the global configuration.

The relay agent relays packets that meet only the following conditions:

- The destination MAC address must be the all-ones broadcast address (FF:FF:FF:FF:FF:FF).
- The destination IP address must be the limited broadcast address (255.255.255.255) or a directed broadcast address for the receive interface.
- The IP time-to-live (TTL) must be greater than 1.

- The protocol field in the IP header must be UDP (17).
- The destination UDP port must match a configured relay entry.

The following table shows the most common protocols and their UDP port numbers and names that are relayed.

UDP Port Number	*Acronym*	*Application*
7	Echo	Echo
11	SysStat	Active User
15	NetStat	NetStat
17	Quote	Quote of the day
19	CHARGEN	Character Generator
20	FTP-data	FTP Data
21	FTP	FTP
37	Time	Time
42	NAMESERVER	Host Name Server
43	NICNAME	Who is
53	DOMAIN	Domain Name Server
69	TFTP	Trivial File Transfer
111	SUNRPC	Sun Microsystems Rpc
123	NTP	Network Time
137	NetBiosNameService	NT Server to Station Connections
138	NetBiosDatagramService	NT Server to Station Connections
139	NetBios	SessionServiceNT Server to Station Connections
161	SNMP	Simple Network Management
162	SNMP-trap	Simple Network Management Traps
513	who	Unix Rwho Daemon
514	syslog	System Log
525	timed	Time Daemon

Table 7-3: UDP Port Allocation

7.4.1. Relay Agent Configuration Example

The example in this section shows how to configure the L3 relay agent (IP helper) to relay and discard various protocols.

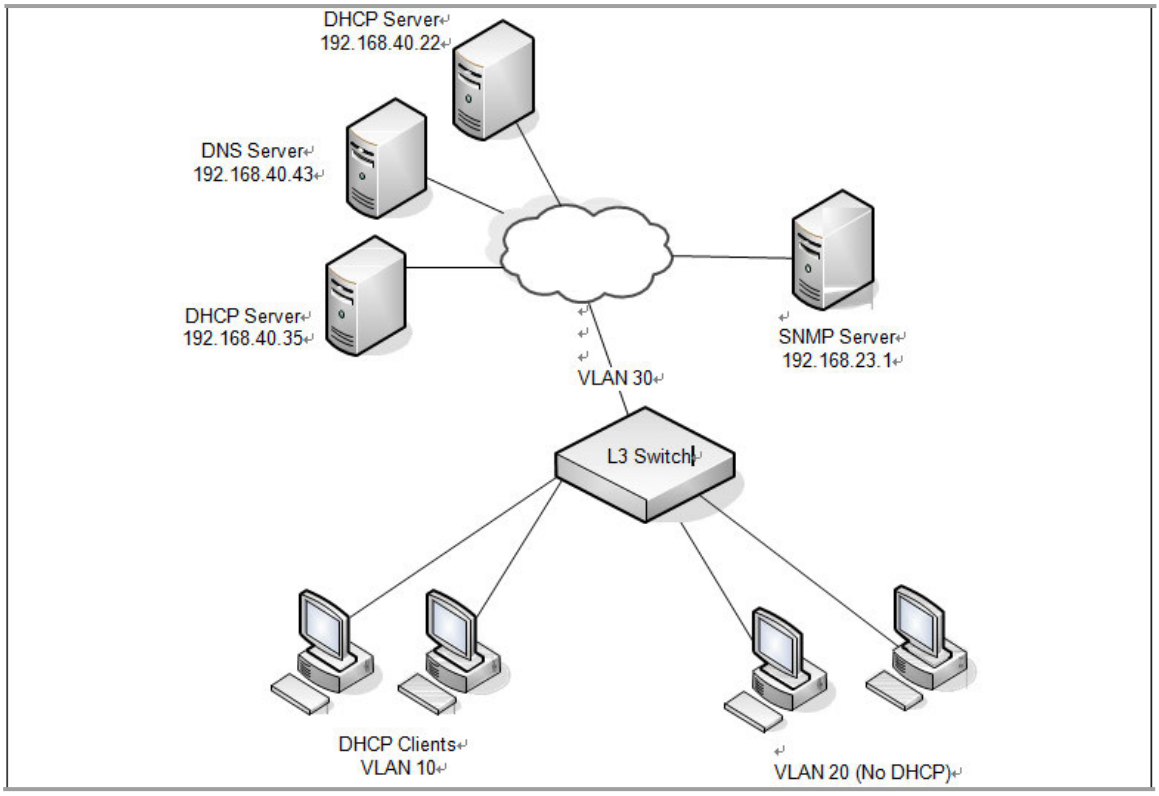


Figure 7-7: L3 Relay Network Diagram

This example assumes that multiple VLAN routing interfaces have been created and configured with IP addresses.

To configure the switch:

1. Enable IP helper on the switch.

```
(Switch) #config
(Switch) (Config)#ip helper enable
```

2. Relay DHCP packets received on VLAN 10 to 192.168.40.35.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip helper-address 192.168.40.35 dhcp
```

3. Relay DNS packets received on VLAN 10 to 192.168.40.43.

```
(Switch) (if-vlan10)#ip helper-address 192.168.40.35 domain
(Switch) (if-vlan10)#exit
```

4. Relay SNMP traps (port 162) received on VLAN 20 to 192.168.23.1.

```
(Switch) (Config)#interface vlan 20
(Switch) (if-vlan20)#ip helper-address 192.168.23.1 162
```

5. The clients on VLAN 20 have statically-configured network information, so the switch is configured to drop DHCP packets received on VLAN 20.

```
(Switch) (if-vlan20)#ip helper-address discard dhcp

(Switch) (if-vlan20)#exit
```

6. Configure the switch so that DHCP packets received from clients in any VLAN other than VLAN 10 and VLAN 20 are relayed to 192.168.40.22.

Note: The following command is issued in Global Configuration mode, so it applies to all interfaces except VLAN 10 and VLAN 20. IP helper commands issued in Interface Configuration mode override the commands issued in Global Configuration Mode.

```
(Switch) (Config)#ip helper-address 192.168.40.22 dhcp

(Switch) (Config)#exit
```

7. Verify the configuration.

```
(Switch) #show ip helper-address


IP helper is enabled


Interface               UDP Port    Discard  Hit Count     Server Address
--------------------- ----------- ---------- ---------- ------------------
vlan 10                   domain          No          0      192.168.40.35
vlan 10                     dhcp          No          0      192.168.40.35
vlan 20                     dhcp         Yes          0
vlan 20                      162          No          0       192.168.23.1
Any                         dhcp          No          0      192.168.40.22
```

7.5. Border Gateway Patrol (BGP)

BGP is an exterior routing protocol that maintains routing tables, transmits routing updates, and bases routing decisions on routing metrics through exchanges of Network Layer Reachability Information (NLRI) with network peers (known as neighbors) via TCP/IP sessions. BGP relies on the local route table, which is populated by IGP routing protocols, in order to establish connectivity for routes contained within NLRI definitions. For routes with established connectivity, BGP determines the best route among those learned from one or more peers and then installs those routes to the local route table as well as advertises those routes to its other peers. Local policy configuration is commonly used to filter NLRIs inbound and outbound, as well as for modifying the attributes of NLRIs that are advertised to peers.

7.5.1. BGP Topology

BGP maintains routing information between routers within different Autonomous Systems (AS), where each AS typically encapsulates a single IGP routing domain. BGP peers exchange NLRIs that contain an AS path, which is an ordered set of AS values that describe the autonomous systems that must be traversed to reach a network destination. Using a distance vector algorithm, BGP uses the AS path to determine the relative distance to a network destination, and detects any potential routing loops. BGP has two types of relationships with its network peers: External BGP peering (EBGP) and Internal BGP peering (IBGP).

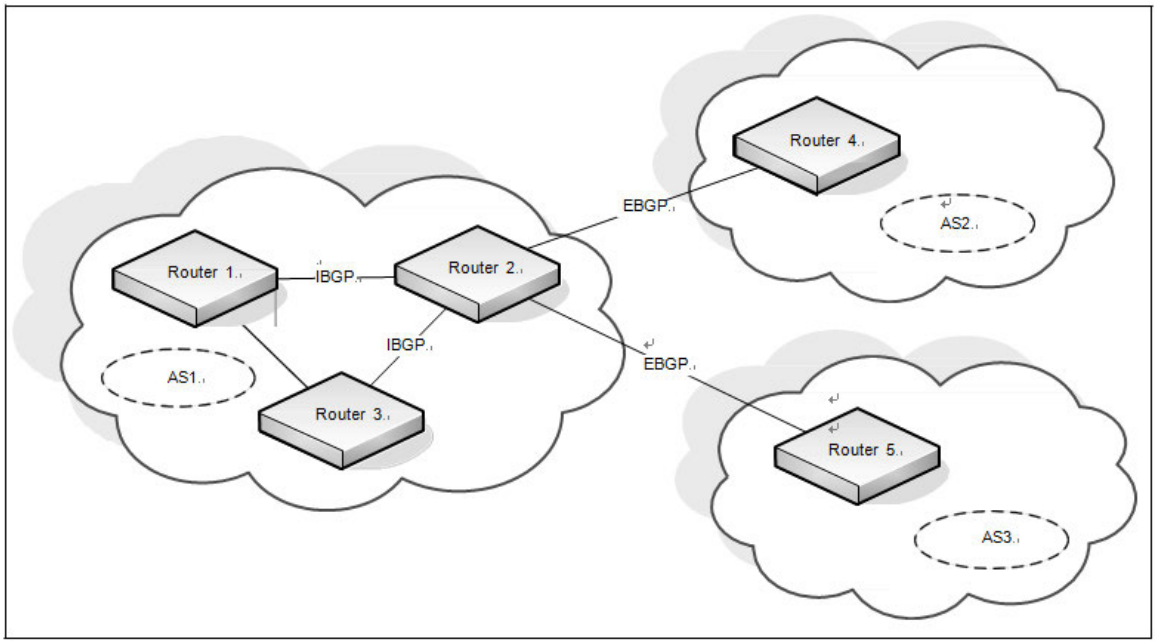


Figure 7-8: Example BGP Network

7.5.1.1. External BGP Peering

EBGP peering occurs between two or more BGP routers in different AS's. Peer routers in these different AS's use BGP to maintain a consistent view of the inter-network topology. External BGP peers exchange NLRIs, which contain reachable network destinations along with BGP specific attributes such as AS path information and various metrics. These BGP attributes along with local policy configuration, which is used to filter and/or modify the BGP NLRIs, are used by BGP to determine optimal routes to these network destinations within the Internet. An illustration of the above scenario can be observed in the previous figure between Router 2 and Router 4.

7.5.1.2. Internal BGP Peering

IBGP peering occurs between two or more BGP routers located within the same AS. Internal BGP peers are mainly responsible for distributing BGP NLRIs, which have been acquired via External BGP peers, to all other Internal BGP peers within the AS. The BGP protocol requires that all IBGP peers within an AS are logically connected as a “full mesh.” Thus, all BGP routers within the AS can have a consistent view of the inter-network destinations. An illustration of the above scenario can be observed in the previous figure between Router R1 and Router 2.

7.5.1.3. Advertising Network Layer Reachability Information

In addition to NLRIs exchanged between BGP peers, a BGP router may originate NLRIs for advertisement to its peers due to local configuration of “locally-originated” routes or “redistribution” policy. In this scenario, the configuration of locally-originated routes or redistribution policy maps to routes installed in the local router's forwarding table by IGP routing protocols on the local router. These routes typically define reachability to network destinations within the local AS. In this manner, BGP is used to advertise NLRIs that define reachability to network destinations within its own AS to BGP peers outside of the local AS.

7.5.2. BGP Behavior

To begin with, BGP systems form a TCP/IP connection between one another to exchange NLRIs. First, they exchange messages to open and confirm the connection parameters. The initial data flow is the entire BGP routing table. Incremental updates are sent as the routing tables change. BGP does not require periodic refresh of the entire BGP routing table because it relies on the reliable transport provided by TCP. Therefore, a BGP speaker must retain the current version of the entire BGP routing tables of all of its peers for the duration of the connection. Keepalive messages are sent periodically to ensure that connection is active. Notification messages are sent in response to errors or special conditions. If a connection encounters an error condition, a notification message is sent and the connection is closed.Routes are advertised between a pair of BGP speakers in UPDATE messages, where the network destinations are the systems whose IP addresses are reported in the NLRI field, and the AS path for those destinations is part of the information reported in the path attributes fields of the same UPDATE message, along with various other BGP attributes. Routes are stored in local Routing Information Bases (RIBs). Logically, all routes learned from a particular BGP peer are kept in a local Adj-RIB-In, and all routes learned from all BGP peers are held in a Loc-RIB, which serves as the central database for BGP to determine the *best* path to a particular network destination. Additionally, local policy configuration may filter or modify the BGP attributes of NLRIs that are received from BGP peers.

Once BGP has chosen the *best* path to a network destination based on the BGP attributes given in an NLRI (also known as the *decision process*), it must determine if there is connectivity to the destination defined by the BGP *nexthop* attribute from the *best* NLRI. Here, BGP performs *nexthop resolution* by referencing the local router's forwarding table, which is populated with routes installed by IGP protocols. If connectivity to the BGP nexthop is found (i.e. resolved), then the corresponding BGP route can be installed to the local router's forwarding table, using the *real* nexthop information from the IGP route that was used to resolve the BGP nexthop.

Finally, BGP routes that have been installed in the local router's forwarding table are eligible to be advertised to connected BGP peers. BGP advertises these routes to each connected peer, typically resetting the BGP nexthop attribute to be the local IP address for the BGP peer connection. Additionally, local policy configuration may filter or modify the NLRIs that are advertised to these BGP peers.

For a more detailed and comprehensive description of BGP protocol behavior, refer to the BGP-4 Protocol Specification (RFC1771/draft-ietf-idr-bgp4-26).

7.5.2.1. BGP Route Selection

On the switch, BGP uses the following route selection rules:

1. Prefer the route with the higher local preference

2. Prefer a locally-originated route over a non-locally originated route

3. Prefer the route with the shorter AS Path

4. Prefer the route with the lower ORIGIN. IGP is better than EGP is better than INCOMPLETE.

5. Prefer the route with the lower MED. By default, MEDs are only compared for routes from the same AS, but a configuration option allows comparison of MEDs from different ASs. A route with no MED is considered to have a MED of 0.

6. Prefer an eBGP route to an iBGP route

7. Prefer the route with the lower IGP cost to the BGP NEXT HOP

8. Prefer the route learned from the peer with the lower router ID

9. Prefer the route learned from the peer with the lower peer IP address

7.5.3. BGP Configuration Example

The following figure shows the topology of a large network that includes two autonomous systems. The commands in this example configure Router 3 (R3) in AS 65049 and Router 9 in AS 65001.

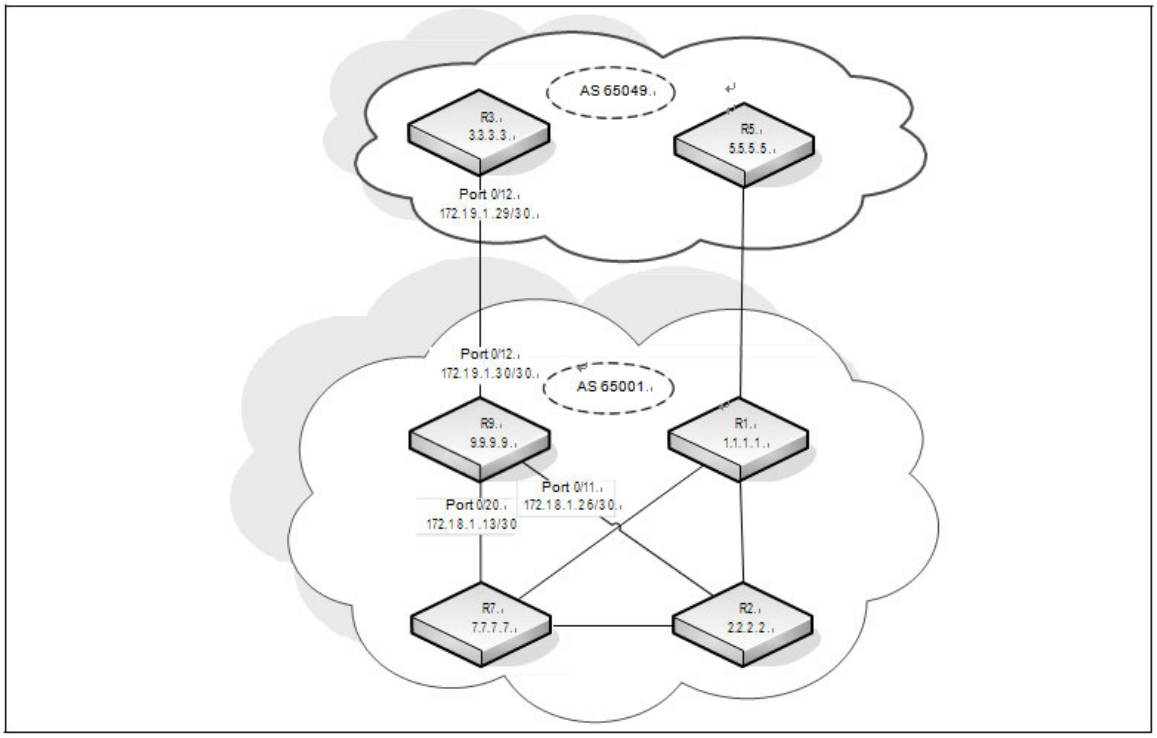


Figure 7-9: BGP Configuration Example

7.5.3.1. Configuring BGP on Router 9

To configure Router 9 (R9) as shown in the previous figure:

1. To make it easier to determine which device is being configured, set the name of Router 9 (R9) as the system prompt.

```
(Switch) #configure
(Switch) (Config)#hostname R9
(R9) (Config)#
```

2. Enter Global Config mode and enable routing on the system.

```
(R9) #configure
(R9) (Config)#ip routing
```

3. Enter Interface Config mode for port 0/11. This interface is connected to R2, which is part of the same AS.

Assign an IP address to the interface, and enable routing on the interface.

```
(R9) (Config)#interface 0/11
(R9) (Interface 0/11)#ip address 172.18.1.26 255.255.255.252
(R9) (Interface 0/11)#routing
```

4. Configure the OSPF timers. The hello interval should be the same on all routers attached to a common network. Likewise, the dead interval timers should be the same across all routers on the network.

```
(R9) (Interface 0/11)#ip ospf hello-interval 3
(R9) (Interface 0/11)#ip ospf dead-interval 12
```

5. Configure OSPF to treat the interface as a point-to-point link.

```
(R9) (Interface 0/11)#ip ospf network point-to-point
(R9) (Interface 0/11)#exit
```

6. Enter Interface Config mode for port 0/12. This is the interface that is connected to R3, which is in a different AS. Assign an IP address to the interface, and enable routing on the interface.

```
(R9) (Interface 0/12)#interface 0/12
(R9) (Interface 0/12)#ip address 172.19.1.30 255.255.255.252
(R9) (Interface 0/12)#routing
(R9) (Interface 0/12)#exit
```

7. Enter Interface Config mode for port 0/20. This interface is connected to R7, which is part of the same AS. Assign an IP address to the interface, and enable routing on the interface.

```
(R9) (Interface 0/20)#interface 0/20
(R9) (Interface 0/20)#ip address 172.18.1.13 255.255.255.252
```

```
(R9) (Interface 0/20)#routing
```

8. Configure the OSPF timers.

```
(R9) (Interface 0/20)#ip ospf hello-interval 3 (R9) (Interface 0/20)#ip ospf dead-interval 12
```

9. Configure OSPF to treat the interface as a point-to-point link.

```
(R9) (Interface 0/20)#ip ospf network point-to-point
(R9) (Interface 0/20)#exit
```

10. Enter Interface Config mode for loopback interface 0 and assign an IP address to the interface.

```
(R9) (Config)#interface loopback 0
(R9) (Interface loopback 0)#ip address 192.168.0.9 255.255.255.255
```

11. Configure the OSPF area ID that the loopback interface belongs to.

```
(R9) (Interface loopback 0)#ip ospf area 0
(R9) (Interface loopback 0)#exit
```

12. Configure the OSPF settings for the router.

```
(R9) (Config)#router ospf
(R9) (Config-router)#router-id 9.9.9.9
(R9) (Config-router)#network 172.19.1.0 0.0.0.255 area 0
(R9) (Config-router)#network 172.18.1.0 0.0.0.255 area 0
(R9) (Config-router)#passive-interface 0/12
(R9) (Config-router)#timers spf 3 5
(R9) (Config-router)#max-metric router-lsa summary-lsa on-startup 90
(R9) (Config-router)#exit
```

13. Enable BGP and identify the autonomous system (AS) number of the router.

```
(R9) (Config-router)#router bgp 65001
```

14. Configure the BGP router ID.

```
(R9) (Config-router)#bgp router-id 9.9.9.9
```

15. Specify the maximum number of next hops BGP may include in an Equal Cost Multipath (ECMP) route derived from paths received from neighbors *outside* the local autonomous system.

```
(R9) (Config-router)#maximum-paths 24
```

16. Set the maximum number of next hops BGP may include in an ECMP route derived from paths received from neighbors *within* the local autonomous system.

```
(R9) (Config-router)#maximum-paths ibgp 24
```

17. Enable the logging of adjacency state changes.

```
(R9) (Config-router)#bgp log-neighbor-changes
```

18. Allow the aggregation of routes with different MED attributes.

```
(R9) (Config-router)#bgp aggregate-different-meds
```

19. Configure the keepalive and hold times that BGP uses for all of its neighbors.

```
(R9) (Config-router)#timers bgp 4 12
```

20. Configure the summary addresses for BGP.

```
(R9) (Config-router)#aggregate-address 172.16.1.0 255.255.255.0 summary-only
(R9) (Config-router)#aggregate-address 172.17.1.0 255.255.255.0 summary-only
(R9) (Config-router)#aggregate-address 172.18.1.0 255.255.255.0 summary-only
(R9) (Config-router)#aggregate-address 172.19.1.0 255.255.255.0 summary-only
```

21. Configure the networks that are attached to AS 65001.

```
(R9) (Config-router)#network 172.18.1.12 mask 255.255.255.252
(R9) (Config-router)#network 172.18.1.16 mask 255.255.255.252
(R9) (Config-router)#network 172.18.1.20 mask 255.255.255.252
(R9) (Config-router)#network 172.18.1.24 mask 255.255.255.252
(R9) (Config-router)#network 172.17.1.4 mask 255.255.255.252
(R9) (Config-router)#network 172.17.1.8 mask 255.255.255.252
(R9) (Config-router)#network 172.17.1.12 mask 255.255.255.252
(R9) (Config-router)#network 172.19.1.28 mask 255.255.255.252
(R9) (Config-router)#network 172.19.1.32 mask 255.255.255.252
```

22. Configure the loopback addresses of routers in AS 65001.

```
(R9) (Config-router)#network 192.168.0.1 mask 255.255.255.255
(R9) (Config-router)#network 192.168.0.2 mask 255.255.255.255
(R9) (Config-router)#network 192.168.0.9 mask 255.255.255.255
(R9) (Config-router)#network 192.168.0.11 mask 255.255.255.255
(R9) (Config-router)#neighbor 192.168.0.11 remote-as 65001
(R9) (Config-router)#neighbor 192.168.0.11 description R7
(R9) (Config-router)#neighbor 192.168.0.11 next-hop-self
(R9) (Config-router)#neighbor 192.168.0.11 update-source loopback 0
(R9) (Config-router)#neighbor 192.168.0.1 remote-as 65001
(R9) (Config-router)#neighbor 192.168.0.1 description R1
(R9) (Config-router)#neighbor 192.168.0.1 next-hop-self
(R9) (Config-router)#neighbor 192.168.0.1 update-source loopback 0
(R9) (Config-router)#neighbor 192.168.0.2 remote-as 65001
(R9) (Config-router)#neighbor 192.168.0.2 description R2
(R9) (Config-router)#neighbor 192.168.0.2 next-hop-self
(R9) (Config-router)#neighbor 192.168.0.2 update-source loopback 0
(R9) (Config-router)#neighbor 172.19.1.29 remote-as 65049
```

```
(R9) (Config-router)#neighbor 172.19.1.29 description R3
(R9) (Config-router)#exit
(R9) (Config)#exit
```

7.5.3.2. Configuring BGP on Router 3

To configure Router 3 (R3) as shown in the prervious figure:

1. To make it easier to determine which device is being configured, set the name of Router 3 (R3) as the system prompt.

```
(Switch) #configure
(Switch) (Config)#hostname R3
(R3) (Config)#
```

2. Enter Global Config mode and enable routing on the system.

```
(R3) #configure
(R3) (Config)#ip routing
```

3. Enter Interface Config mode for port 0/12. This is the interface that is connected to R3, which is in a different AS. Assign an IP address to the interface, and enable routing on the interface.

```
(R3) (Interface 0/12)#interface 0/12
(R3) (Interface 0/12)#ip address 172.19.1.29 255.255.255.252
(R3) (Interface 0/12)#routing
(R3) (Interface 0/12)#exit
```

4. Enter Interface Config mode for loopback interface 0 and assign an IP address to the interface.

```
(R3) (Config)#interface loopback 0
(R3) (Interface loopback 0)#ip address 192.168.2.3 255.255.255.255
(R3) (Interface loopback 0)#exit
```

5. Enable BGP and identify the autonomous system (AS) number of the router.

```
(R3) (Config-router)#router bgp 65049
```

6. Configure the BGP router ID.

```
(R3) (Config-router)#bgp router-id 3.3.3.3
```

7. Specify the maximum number of next hops BGP may include in an ECMP route derived from paths received from neighbors outside the local autonomous system.

```
(R3) (Config-router)#maximum-paths 4
```

8. Enable the logging of adjacency state changes.

```
(R3) (Config-router)#bgp log-neighbor-changes
```

9. Configure BGP to advertise connected routes with a metric value of 100.

```
(R3) (Config-router)#redistribute connected metric 100
```

10. Configure the keepalive and hold times that BGP uses for all of its neighbors.

```
(R3) (Config-router)#timers bgp 4 12
```

11. Configure the loopback addresses of routers in AS 65049.

```
(R3) (Config-router)#network 192.168.2.3 mask 255.255.255.255
(R3) (Config-router)#neighbor 172.19.1.30 remote-as 65001
(R3) (Config-router)#neighbor 172.19.1.30 description R9
(R3) (Config-router)#exit
(R3) (Config)#exit
```

7.6. IPv6 Routing

IPv6 is the next generation of the Internet Protocol. With 128-bit addresses, versus 32-bit addresses for IPv4, IPv6 solves the address depletion issues seen with IPv4 and removes the requirement for Network Address Translation (NAT), which is used in IPv4 networks to reduce the number of globally unique IP addresses required for a given network.

On the switch, IPv6 coexists with IPv4. As with IPv4, IPv6 routing can be enabled on loopback and VLAN interfaces. Each L3 routing interface can be used for IPv4, IPv6, or both. IP protocols running over L3 (for example, UDP and TCP) are common to both IPv4 and IPv6.

7.6.1. How Does IPv6 Compare with IPv6

There are many conceptual similarities between IPv4 and IPv6 network operation. Addresses still have a network prefix portion (network) and a device interface specific portion (host). While the length of the network portion is still variable, most users have standardized on using a network prefix length of 64 bits. This leaves 64 bits for the interface specific portion, called an Interface ID in IPv6. Depending upon the underlying link addressing, the Interface ID can be automatically computed from the link (e.g., MAC address). Such an automatically computed Interface ID is called an EUI-64 identifier, which is the interface MAC address with ff:fe inserted in the middle.

IPv6 packets on the network are of an entirely different format than traditional IPv4 packets and are also encapsulated in a different EtherType (86DD rather than 0800 which is used with IPv4). The details for encapsulating IPv6 in Ethernet frames are described in RFC4862.

Unlike IPv4, IPv6 does not have broadcasts. There are two types of IPv6 addresses — unicast and multicast. Unicast addresses allow direct one-to-one communication between two hosts, whereas multicast addresses allow one-to-many communication. Multicast addresses are used as destinations only. Unicast addresses will have 00 through fe in the most significant octets and multicast addresses will have ff in the most significant octets.

7.6.2. How are IPv6 Interface Configured

On the switch, IPv6 can coexists with IPv4. As with IPv4, IPv6 routing can be enabled on VLAN interfaces. Each L3 routing interface can be used for IPv4, IPv6, or both simultaneously.

Neighbor Discovery (ND) protocol is the IPv6 replacement for Address Resolution Protocol (ARP) in IPv4. The IPv6 Neighbor Discovery protocol is described in detail in RFC4861. Router advertisement is part of the Neighbor Discovery process and is required for IPv6. As part of router advertisement, the switch supports stateless auto configuration of end nodes. The switch supports both EUI-64 interface identifiers and manually configured interface IDs.

While optional in IPv4, router advertisement is mandatory in IPv6. Router advertisements specify the network prefix or prefixes on a link which can be used by receiving hosts, in conjunction with an EUI-64 identifier, to autoconfigure a host's address. Routers have their network prefixes configured and may use EUI-64 or manually configured interface IDs.

In addition to zero or more global addresses, each IPv6 interface also has an autoconfigured "link-local" address which is:

- fe80::/10, with the EUI-64 address in the least significant bits.
- Reachable only on the local VLAN — link-local addresses are never routed.
- Not globally unique

Next hop addresses computed by routing protocols are usually link-local addresses.

During the period of transitioning the Internet to IPv6, a global IPv6 Internet backbone may not be available. One transition mechanism is to tunnel IPv6 packets inside IPv4 to reach remote IPv6 islands. When a packet is sent over such a link, it is encapsulated in IPv4 in order to traverse an IPv4 network and has the IPv4 headers removed at the other end of the tunnel.

7.6.3. Default IPv6 Routing Values

The following table shows the default values for the IP routing features this section describes.

Parameter	***Default Value***
IPv6 Unicast Routing Mode	Disabled
IPv6 Hop Limit	Unconfigured
ICMPv6 Rate Limit Error Interval	1000 milliseconds
ICMPv6 Rate Limit Burst Size	100 messages
Interface IPv6 Mode	Disabled
IPv6 Router Route Preferences	Local—0 Static—1 OSPFv3 Intra—110 OSPFv3 Inter—110 OSPFv3 External—110 BGP External—20 BGP Internal—200 BGP Local —200

Table 7-4: IPv6 Routing Defaults

The following table shows the default IPv6 interface values after a VLAN routing interface has been created.

Parameter	*Default Value*
IPv6 Mode	Disabled
DHCPv6 Client Mode	Disabled
Stateless Address AutoConfig Mode	Disabled
Routing Mode	Enabled
Interface Maximum Transmit Unit	1500
Router Duplicate Address Detection Transmits	1
Router Advertisement NS Interval	Not configured
Router Lifetime Interval	1800 seconds
Router Advertisement Reachable Time	0 seconds
Router Advertisement Interval	600 seconds
Router Advertisement Managed Config Flag	Disabled
Router Advertisement Other Config Flag	Disabled
Router Advertisement Suppress Flag	Disabled
IPv6 Destination Unreachables	Enabled

Table 7-5: IPv6 Interface Defaults

7.6.4. Configuring IPv6 Routing Features

This section provides information about the commands you use to configure IPv6 routing on the switch.

7.6.4.1. Configuring Global IP Routing Settings

Use the following commands to configure various global IP routing settings.

Command	*Purpose*
configure	Enter global configuration mode.
sdm prefer dual-ipv4-and-ipv6 {alpm | data-center | dcvpn-data-center | default}	Select a Switch Database Management (SDM) template to enable support for both IPv4 and IPv6. Changing the SDM template requires a system reload.
ipv6 unicast-routing	Golbally enable IPv6 routing on the switch.
ipv6 hop-limit <hops>	Set the TTL value for the router. The valid range is 0 to 255.
ipv6 icmp error-interval <burst-interval> [<burst-size>]	Limit the rate at which IPv6 ICMP error messages are sent. . `burst-interval` – How often the token bucket is initialized (Range: 0-2147483647 milliseconds).
exit	Exit to Privileged EXEC mode.
show ipv6 protocols	This command lists a summary of the configuration and status for each of the active IPv6 routing protocols. If a protocol is selected on the command line, the display will be limited to that protocol.

Table 7-6: Global IP Routing Settings

7.6.4.2. Configuring IPv6 Interface Settings

Use the following commands to configure IPv6 settings for VLAN, tunnel, or loopback interfaces.

Command	Purpose
configure	Enter Global Configuration mode.
interface {vlan | tunnel | loopback} interface-id	Enter Interface Configuration mode for the specified VLAN, tunnel, or loopback interface.
ipv6 enable	Enable IPv6 on the interface. Configuring an IPv6 address will automatically enable IPv6 on the interface.
ipv6 address {autoconfig | dhcp | prefix/prefix-length [eui64]}	Configure the IPv6 address and network prefix length. Setting an IPv6 address enables IPv6 on the interface. You can also use the **ipv6 enable command** to enable IPv6 on the interface without setting an address. Link-local, multicast, IPv4-compatible, and IPv4-mapped addresses are not allowed to be configured. Include the **EUI-64** keyword to have the system add the 64-bit interface ID to the address. You must use a network prefix length of 64 in this case. For VLAN interfaces, use the **dhcp** keyword to enable the DHCPv6 client and obtain an IP address form a network DHCPv6 server.
ipv6 mtu	(VLAN interfaces only) Set the IPv6 Maximum Transmission Unit (MTU) on a routing interface. The IPv6 MTU is the size of the largest IPv6 packet that can be transmitted on the interface without fragmentation. The range is 1280–12270 bytes.
ipv6 traffic-filter ACL name	Add an access-list filter to this interface.
ipv6 unreachables	Allow the interface to send ICMPv6 Destination Unreachable messages. The no ipv6 unreachables command suppresses the ICMPv6 unreachable messages for this interface.
exit	Exit the interface configuration mode.

Table 7-7: IPv6 Interface Settings

7.6.4.3. Configuring IPv6 Neighbor Discovery

Use the following commands to configure IPv6 Neighbor Discovery settings.

Command	Purpose
ipv6 nd prefix prefix/ prefix-length [{valid-lifetime| infinite} {preferred-lifetime| infinite}] [no-autoconfig] [off-link]	Configure parameters associated with network prefixes that the router advertises in its Neighbor Discovery advertisements. • **ipv6-prefix**—IPv6 network prefix. • ***prefix-length***—IPv6 network prefix length. • ***valid-lifetime***—Valid lifetime of the router in seconds. (Range: 0–4294967295 seconds.) • **infinite**—Indicates lifetime value is infinite. • ***preferred-lifetime***—Preferred-lifetime of the router in seconds. (Range: 0–4294967295 seconds.) • **no-autoconfig**—Do not use the prefix for auto configuration. • **off-link**—Do not use the prefix for onlink determination.
ipv6 nd ra-interval maximum minimum	Set the transmission interval between router Neighbor Discovery advertisements. • ***maximum*** — The maximum interval duration (Range: 4–1800 seconds). • ***minimum*** — The minimum interval duration (Range: 3 – (0.75 * maximum) seconds).
ipv6 nd ra-lifetime seconds	• Set the value that is placed in the Router Lifetime field of the router Neighbor Discovery advertisements sent from the interface. • The *seconds* value must be zero, or it must be an integer between the value of the router advertisement transmission interval and 9000 seconds. A value of zero means this router is not to be used as the default router. (Range: 0-9000).
ipv6 nd suppress-ra	• Suppress router advertisement transmission on an interface.
ipv6 nd dad attempts value	• Set the number of duplicate address detection probes transmitted while doing Neighbor Discovery. • The range for *value* is 0–600.
ipv6 nd ns-interval milliseconds	• Set the interval between router advertisements for advertised neighbor solicitations. The range is 1000 to 4294967295 milliseconds.
ipv6 nd other-config-flag	• Set the *other stateful configuration* flag in router advertisements sent from the interface.
ipv6 nd managed-config-flag	• Set the *managed address configuration* flag in router advertisements. When the value is true, end nodes use DHCPv6. When the value is false, end nodes automatically configure addresses.
ipv6 nd reachable-time milliseconds	• Set the router advertisement time to consider a neighbor reachable after neighbor discovery confirmation.
ipv6 neighbors dynamicrenew	• Enables/disables the periodic NUD (neighbor unreachability detection) to be run on the existing IPv6 neighbor entries based on the activity of the entries in the hardware. If the setting is disabled, only those entries that are actively used in the hardware are triggered for NUD at the end of STALE timeout of 1200 seconds. If the setting is enabled, periodically every 40 seconds a set of 300 entries are triggered for NUD irrespective of their usage in the hardware.
ipv6 nud max-unicast-solicits	• Configures the maximum number of unicast Neighbor Solicitations sent during neighbor resolution or during NUD (neighbor unreachability detection). The value ranges from 3 to 10.
ipv6 nud max-multicast-solicits	• Configures the maximum number of multicast Neighbor Solicitations sent during neighbor resolution or during NUD (neighbor unreachability detection). The value ranges from 3 to 255.
ipv6 nud backoff-multiple	• Configures the exponential backoff multiple to be used in the calculation of the next timeout value for Neighbor Solicitation transmission during NUD (neighbor unreachability detection) following the exponential backoff algorithm. The value ranges from 1 to 5. The next timeout value is limited to a maximum value of 60 seconds if the value with exponential backoff calculation is greater than 60 seconds.

Table 7-8: IPv6 Neighbor Discovery Settings

7.6.4.4. Configuring IPv6 Route Table Entries and Route Preferences

Use the following commands to configure IPv6 Static Routes.

Command	Purpose
configure	Enter global configuration mode.
ipv6 route ipv6-prefix/prefix-length {next-hop-address | interface-type interface-number next-hop-address } [preference]	Configure a static route. Use the keyword **null** instead of the next hop router IP address to configure a static reject route. • ***prefix/prefix-length***—The IPv6 network prefix and prefix length that is the destination of the static route. Use the ::/0 form (unspecified address and zero length prefix) to specify a default route. • ***interface-type interface-number***—Must be specified when using a link-local address as the next hop. The interface-type can be **vlan** or **tunnel**. • ***next-hop-address*** —The IPv6 address of the next hop that can be used to reach the specified network. A link-local next hop address must have a prefix length of 128. The next hop address cannot be an unspecified address (all zeros), a multicast address, or a loopback address. If a link local next hop address is specified, the interface (VLAN or tunnel), must also be specified. • ***preference***—Also known as Administrative Distance, a metric the router uses to compare this route with routes from other route sources that have the same network prefix. (Range: 1-255). Lower values have precedence over higher values. The default preference for static routes is 1. Routes with a preference of 255 are considered as "disabled" and will not be used for forwarding. Routes with a preference metric of 254 are used by the local router but will never be advertised to other neighboring routers.
ipv6 route ipv6-prefix/prefix-length null [preference]	Configure a static reject route. IPv6 packets matching the reject route will be silently discarded.
ipv6 route distance integer	Set the default distance (preference) for static IPv6 routes. Lower route preference values are preferred when determining the best route. The default distance (preference) for static routes is 1.
exit	Exit to Global Config mode.
serviceport ipv6 neighbor ipv6_neighbor mac_address	Configures a static IPv6 neighbor with the given IPv6 address and MAC address on the service port.
network ipv6 neighbor ipv6_neighbor mac_address	Configures a static IPv6 neighbor with the given IPv6 address and MAC address on the network port.
ipv6 neighbor ipv6_neighbor if_name mac_address	Configures a static IPv6 neighbor if_name with the given IPv6 address and MAC address on the network port.
show serviceport ipv6 neighbors	This command displays the information about the IPv6 neighbor entries cached on the service port. The information is updated to show the type of the entry.
show network ipv6 neighbors	This command displays the information about the IPv6 neighbor entries cached on the network port. The information is updated to show the type of the entry.

Table 7-9: IPv6 Static Routes

7.6.4.5. IPv6 Show Commands

Use the following commands to view IPv6 configuration status and related data.

Command	Purpose
`show sdm prefer`	Show the currently active SDM template.
`show sdm prefer dual-ipv4-and-ipv6 {date-center | default}`	Show parameters for the SDM template.
`show ipv6 dhcp interface vlan vlan-id`	View information about the DHCPv6 lease acquired by the specified interface.
`show ipv6 interface {vlan | tunnel | loopback} interface-id`	View the IP interface configuration information for the specified IPv6 routing interface.
`show ipv6 brief`	View the global IPv6 settings for the switch.
`show ipv6 route [ipv6-address | ipv6-prefix/prefix-length | protocol | interface-type interface-number] [best]`	View the routing table. • ***ipv6-address***—Specifies an IPv6 address for which the best-matching route would be displayed. • ***protocol***—Specifies the protocol that installed the routes. Is one of the following keywords: **connected**, **ospf**, **static**. • ***ipv6-prefix/ prefix-length***—Specifies an IPv6 network for which the matching route would be displayed. • ***interface-type interface-number***—Valid IPv6 interface. Specifies that the routes with next hops on the selected interface be displayed. • **best**—Specifies that only the best routes are displayed. If the connected keyword is selected for protocol, the best option is not available because there are no best or non-best connected routes.
`show ipv6 route summary`	View summary information about the IPv6 routing table.
`show ipv6 route preferences`	View detailed information about the IPv6 route preferences.

Table 7-10: IPv6 Configuration Status

7.7. ECMP Hash Selection

Users can choose the load balancing/sharing algorithm used for selecting the final ECMP route. The management interfaces enable choosing various combinations of IP header fields, including the inner or outer IP headers in tunneled packets. Both IPv4 and IPv6 are supported. The field selectors remain the same for all packet types.

- Source IP address of the packet.
- Destination IP address of the packet.
- Source and Destination IP address of the packet.
- Source IP address and Source TCP/UDP Port field associated with the packet.
- Destination IP address and Destination TCP/UDP Port field associated with the packet.
- Source, Destination IP address and Source, Destination TCP/UDP Port field associated with the packet.

For tunneled packets, the user also must select whether the inner or the outer IP header should be used.

For configuration information, see the **ip load-sharing** command in the *NETGEAR M4500 Series Switches CLI Command Reference Manual*.

7.8. Bidirectional Forwarding Detection

In a network device, Bidirectional Forwarding Detection (BFD) is presented as a service to its user applications, providing them options to create and destroy a session with a peer device and reporting upon the session status. On the M4500 series switches, OSPF and BGP can use BFD for monitoring of their neighbors' availability in the network and for fast detection of connection faults with them.

BFD uses a simple 'hello' mechanism that is similar to the neighbor detection components of some well-known protocols. It establishes an operational session between a pair of network devices to detect a two-way communication path between them and serves information regarding it to the user applications. The pair of devices transmits BFD packets between them periodically, and if one stops receiving peer packets within detection time limit it considers the bidirectional path to have failed. It then notifies the application protocol using its services.

BFD allows each device to estimate how quickly it can send and receive BFD packets to agree with its neighbor upon how fast detection of failure could be done.

BFD can operate between two devices on top of any underlying data protocol (network layer, link layer, tunnels, etc.) as payload of any encapsulating protocol appropriate for the transmission medium. The implementation works with IPv4 and IPv6 networks and supports IPv4/v6 address-based encapsulations.

7.8.1. Configuring BFD

The following command sequence enables BFD and configures session parameters:

1. First, globally enable BFD:

```
(Switch)#configure
(Switch) (Config)# feature bfd
```

2. Configure session settings. These can be configured globally or on a per-interface basis.

```
(Switch) (Config)#bfd interval 100 min_rx 200 multiplier 5
(Switch) (Config)#bfd slow-timer 1000
```

- The argument **interval** refers to the desired minimum transmit interval, the minimum interval that the user wants to use while transmitting BFD control packets (in ms).
- The argument **min_rx** refers to the required minimum receive interval, the minimum interval at which the system can receive BFD control packets (in ms).
- The argument **multiplier** specifies the number of BFD control packets to be missed in a row to declare a session down.
- The **slow-timer** command sets up the BFD required echo receive interval preference value (in ms). This value determines the interval the asynchronous sessions use for BFD control packets when the echo

function is enabled. The slow-timer value is used as the new control packet interval, while the echo packets use the configured BFD intervals.

3. Configure BGP to use BFD for fast detection of faults between neighboring devices.

```
(Switch) (Config)#router bgp
(Switch) (Config-router)# neighbor 172.16.11.6 fall-over bfd
(Switch) (Config-router)# exit
```

4. Enable BFD globally for OSPF:

```
(Switch) (Config)#router ospf
(Switch) (Config-router)# bfd
(Switch) (Config-router)# exit
```

5. Configure OSPF to use BFD on the interface:

```
(Switch) #configure
(Switch) (Config)#interface 0/9
(Switch) (Interface 0/9)#ip ospf bfd
(Switch) (Interface 0/9)#exit
```

6. Create static Route to specific destination:

```
(Switch) (Config)# ip route 6.6.6.6 255.255.255.255 155.1.56.6
```

7. Configure Static route to use BFD on the interface:

```
(Switch) (Config)# ip route static bfd 155.1.56.6 155.1.56.5
```

7.9. VRF Lite Operation and Configuration

Each virtual router behaves like an independent router. Virtual routers can be created and destroyed dynamically. The fault domains of virtual routers are isolated. Bringing down a virtual router does not impact another virtual router. Each virtual router has its own instances of routing protocols and routing applications. The total number of routes or host entries is still limited by the hardware capacities on the physical router, but the routes and host entries are distributed across the virtual routing domains based on the user configuration.

IP prefixes can overlap between two VR instances. The same IP address can be configured on two interfaces that are a part of different VR instances. A packet is routed based on the route table look up result in the corresponding VR instance. The VR instance is derived based on the ingress interface. There are situations, however, that require support for inter-VR routing, such as providing access to shared services syslog server, DHCP server, the Internet, etc. These cases are handled through “route leaking”.

By default, all the standard routing software and functions are in the default router (VRID 0), which is created on startup and cannot be deleted by the user. The non-VRF routing user does not experience any disruption in using the CLI commands or in router functionality as a result of VRF configuration.

7.9.1. Route Leaking

Route leaking is the ability to install a route in one VRF that allows traffic to flow to another VRF. Although this mechanism breaks the isolation between VRFs, it is sometimes used to provide access to common services for devices inside the different VRFs. The switch supports route leaking between the global default routing table and a VR, but not across VRs. The switch supports route leaking only through static routes. The switch does not support inter- VRF packet forwarding by connecting a wire between ports belonging to different VR instances

7.9.2. Adding Leaked Routes

Connected routes in one router that are leaked into another VR are referred to as leaked host routes. To add leaked host routes, specify the next-hop interface but not the next-hop address. For leaked routes that are not directly connected (static or dynamic routes), the next-hop address must to be specified in addition to the next- hop interface. The next-hop interface is specified to identify the outgoing VR interface. If the next-hop interface is unspecified, the route is treated as an internal route to the VR.

Internal routes within a router that are added with only a next-hop interface value (and no next-hop address value) are supported only over unnumbered interfaces.

7.9.3. Using Leaked Routes

The line rate forwarding continues to work the same for leaked route destinations in a router as for the internal routes in the router. For bidirectional traffic to work between VRs using leaked routes, the corresponding routes should be leaked between the VRs.

7.9.4. CPU-Originated Traffic

For CPU-originated traffic from different applications (ping, traceroute, syslog, IP helper) that may use the leaked routes to access the destination or shared service, the following conditions are required to ensure proper operation:

1. The source IP address in the originated packets must be mentioned with the source IP option (e.g., ping with source option).

2. In the router where the CPU traffic originates, the route for the source option matching network must be leaked into the virtual router where the next-hop belongs so that the return traffic is directed to the traffic-originating router.

7.9.5. VRF Features Support

The following table lists features and details how they are supported by VRF Lite:

Feature	VRF Support
Network Management	Network management includes the ability to manage the switch via CLI and SNMP. Network management is supported only via the default router. Administrators cannot log into the switch and manage the switch via one of the IP addresses on the non-default VR. The Service Port and the Network Port are always associated with the default router, so the customers are able to manage the switch via these interfaces.
SNMP Management	Only the default router can be managed via SNMP.
AAA	The Authentication, Authorization, and Accounting protocols include services such as the RADIUS client and the TACACS+ client. The switch supports these services only on the default router.
Network Services	The Ping and the Trace Route clients are supported in the Virtual Router context. Other protocols are supported only in the default router. These include the SNTP client, DNS client, sFlow, RPCAP, and Auto Install.
Loopback and Tunnel Interfaces	Loopback interfaces with IPv4 prefixes are supported in the Virtual Router. Loopback interfaces with IPv6 addresses can be configured only in the default router. The number of Loopback interfaces in builds containing the VRF package is increased to 64. The loopback interfaces are shared across VR instances in the system and there is no restriction on the maximum supported per VR. Tunnel interfaces are not supported in the Virtual Router.
IP unnumbered interfaces	IP unnumbered interface cannot be part of non-default VRF instance. This feature is supported only in the default router.
OSPFv2	The OSPFv2 protocol is supported in the Virtual Router. As of the current release, a crash in the OSPFv2 protocol does not cause the switch to reboot. All OSPF features including graceful restart and NSF are supported for OSPFv2 in each VR instance.
OSPF v3	The OSPFv3 protocol is supported only in the default router.
RIP	RIP is not currently supported in the Virtual Router.
VRRP	The Virtual Routing Redundancy Protocol is a fault-tolerance feature that enables two or more routers to appear as one router to the IP clients. If one of the VRRP

	routers fails, another router can take over the data forwarding with minimum interruption to client traffic. The VRRP protocol is supported in the Virtual Router context. The VRRP protocol enables two or more virtual routers running on different physical switches to form a VRRP group. The Virtual Routers running on the same physical switch cannot form a VRRP group with each other.
BGP	The Border Gateway Protocol is intended to be used by the Customer Edge (CE) switch to communicate with other CE switches and PE switches across the Provider Network. This typical VRF-Lite deployment is described in “VRF Lite Deployment Scenarios” on page 274. The BGP protocol runs in the Default Router context and is aware of the Virtual Routers. BGP is used to: 1. Redistribute VPN routes from Virtual Routers on the CE switch to the attached PE in the Provider Network. 2. Leak routes dynamically between different Virtual Routers on the same physical switch. This requires support for BGP extended communities and route targets. In the current implementation, BGP does not support either of the above mentioned functionalities.
IPv6	The current release supports VRF-Lite only for IPv4. IPv6 data forwarding and protocols are not currently supported.
IP Multicast	The current Virtual Routing release supports only IPv4 unicast routing.
Policy Based Routing	PBR is a routing policy feature useful in overriding routing decisions with programmable rules. PBR is supported only in the default router in the current release.
DHCP Server	DHCP Server is not VR-aware in the current release.
DHCP Snooping	The IP Source Guard (IPSG) feature uses DHCP snooping to allow only packets from known sources. IPSG uses DHCP Snooping to snoop the DHCP addresses allocated to connected hosts. The tuple (IP, MAC, VLAN, Interface) uniquely identifies a host. DHCP Snooping is a layer-2 feature and is VRF-agnostic. It works in layer-2 of any VLAN irrespective of whether it belongs to a default router or any virtual router. It applies to all protocols working at L2.
IP Helper	IP Helper relays the broadcast packets received on a Routing interface in the VRF context to the configured server address. The server is looked up in the RTO specific to that VR only. Relay across VRs is not supported.
OpEN API	The applications using existing OpEN APIs are not affected by the VRF feature.

Layer-2 Features	The VRF feature does not affect the switch layer-2 features such as virtual port channels (VPC). However, if VPC is planned to be used on VRF-enabled switches, the VPC ports need to be configured to be in the same routing domain.

Table 7-11: VRF features list

7.9.6. VRF Lite Development Scenarios

The following are two likely deployment scenarios for the VRF-Lite solution:

1. In the Customer edge (CE) devices that interface with the PE (Provider edge) device in the service provider backbone network to provide VPN connectivity for the Enterprise network sites spread across different geographical locations across the internet backbone.

In this scenario, the BGP protocol must be running on the device to support feature extensions required to support:

a. Dynamic route leaking locally between the VRFs to leak the routes to shared services using Route Targets.

b. Exchange the VPN related route information per VR with PE device using extended communities

2. The internal Routers in the Enterprise networks to provide isolation of different departments/offices at layer-3 or routing domain.

This scenario does not mandate that the BGP protocol be running on the device. It can still be run in this scenario to achieve dynamic route leaking only. The IGP protocol (OSPF or RIP) running in the VR instance communicates route information with corresponding peers in the same VR on other CE devices or internal Routers.

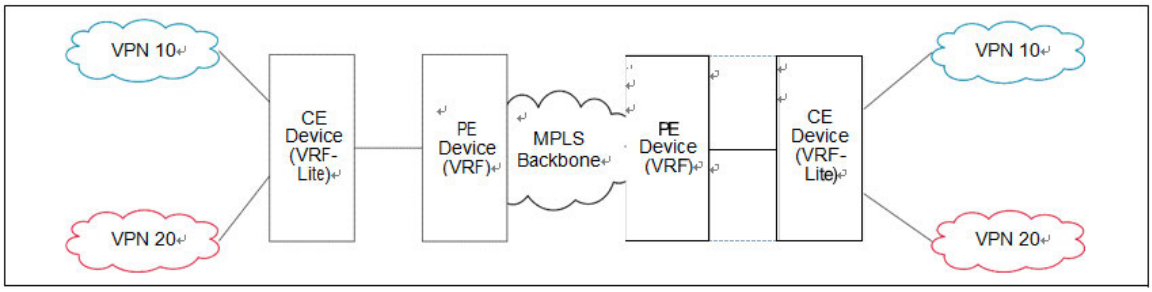


Figure 7-10: VRF Scenarios

The default global routing table is also referred to as VR 0.

In the following example, subnetworks 10.10.10.0/24 and 11.11.11.0/24 belong to the virtual routing domain “HR Dept” and subnetworks 20.20.20.0/24 and 22.22.22.0/24 belong to virtual routing domain “Finance Dept”. Hence, the hosts in networks 10.10.10.0/24 can communicate only with other network 11.11.11.0/24 via the router and the hosts in networks 20.20.20.0/24 can communicate only with other network 22.22.22.0/24 via the router.

If there is a shared service printer @30.30.30.30 in the default global routing domain “Shared Services”, we would want the HR and Finance domains to have access to it. Therefore, we statically leak a 30.30.30.0/24 route

from global routing table to VR 10 and VR 20. At the same time, we statically leak the routes 10.10.10.0/24 and 11.11.11.0/24 from VR 10 to global table (the same applies to VR 20).

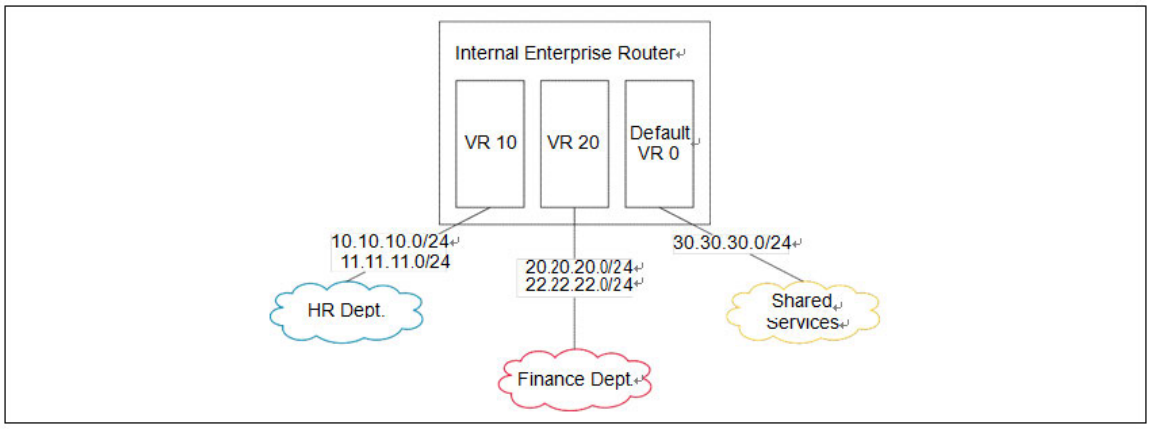


Figure 7-11: VRF routing with shared services

The route tables in both the VRs and the global domain look like the following:

```
(Switch) #show ip route vrf HR
Route Codes: R - RIP Derived, O - OSPF Derived, C - Connected, S - Static
B - BGP Derived, IA - OSPF Inter Area
E1 - OSPF External Type 1, E2 - OSPF External Type 2
N1 - OSPF NSSA External Type 1, N2 - OSPF NSSA External Type 2
L - Leaked Route
C10.10.10.0/24 [0/1] directly connected,vlan 10
C11.11.11.0/24 [0/1] directly connected,vlan 11
S L30.30.30.0/24 [1/1] directly connected,  vlan 30
S L50.50.50.0/24 [1/1] via 30.30.30.2,02d:22h:15m,  vlan 30


(Switch) #show ip route vrf Finance
Route Codes: R - RIP Derived, O - OSPF Derived, C - Connected, S - Static
B - BGP Derived, IA - OSPF Inter Area
E1 - OSPF External Type 1, E2 - OSPF External Type 2
N1 - OSPF NSSA External Type 1, N2 - OSPF NSSA External Type 2
L - Leaked Route


C      20.20.20.0/24 [0/1] directly connected,    vlan 20
C      22.22.22.0/24 [0/1] directly connected,    vlan 22
S L    30.30.30.0/24 [1/1] directly connected,   vlan 30
S L    50.50.50.0/24 [1/1] via 30.30.30.2,       Vlan 30
       02d:22h:15m,
```

```
(Switch) #show ip route

Route Codes: R - RIP Derived, O - OSPF Derived, C - Connected, S - Static

B - BGP Derived, IA - OSPF Inter Area

E1 - OSPF External Type 1, E2 - OSPF External Type 2

N1 - OSPF NSSA External Type 1, N2 - OSPF NSSA External Type 2

L - Leaked Route


C       30.30.30.0/24 [0/1] directly connected,    vlan 30
S L     10.10.10.0/24 [1/1] directly connected,   vlan 10
S L     11.11.11.0/24 [1/1] directly connected,   vlan 11
S L     20.20.20.0/24 [1/1] directly connected,   vlan 20
S L     22.22.22.0/24 [1/1] directly connected,   vlan 22
```

7.9.7. VRF Configuration Example

1. Create virtual router instances. The following commands create and name two instances and enter VRF Configuration mode for each.

In VRF Configuration mode for each VR, a description is added and the maximum number of routes allowed in each virtual instance is configured. On the “Red” instance, the number of routes above which a warning message is issued is also configured.

The **ip routing** command enables routing in each VR instance:

```
(Switch) #configure

(Switch) (Config)#ip vrf Red

(Switch) (Config-vrf-Red)#description "finance department"

(Switch) (Config-vrf-Red)#maximum routes 2048

(Switch) (Config-vrf-Red)#maximum routes warn 80

(Switch) (Config-vrf-Red)#ip routing

(Switch) (Config-vrf-Red)#exit


(Switch) (Config)#ip vrf Blue

(Switch) (Config-vrf-Blue)#description "human resources department"

(Switch) (Config-vrf-Blue)#maximum routes 4096

(Switch) (Config-vrf-Blue)#ip routing

(Switch) (Config-vrf-Blue)#exit
```

2. In Interface Config mode, assign interfaces to each virtual router:

```
(Switch) (Config)#interface 0/1
```

```
(Switch) (Interface 1/0/1)#ip vrf forwarding Red

Warning: routing interface moved from Default router instance to "Red" router instance. (Switch)
(Interface 1/0/1)#exit


(Switch) (Config)#interface 0/2

(Switch) (Interface 1/0/2)#ip vrf forwarding Blue

Warning: routing interface moved from Default router instance to "Blue" router instance.
(Switch) (Interface 1/0/2)#exit
```

3. Create static leaked routes as needed in the VR instances.

In the following example, subnetwork 9.0.0.0/24 is a connected subnetwork in the global route table and subnet 56.6.6.0/24 is reachable via a gateway 9.0.0.2 in the global route table. Subnet 8.0.0.0/24 is a connected subnetwork in virtual router Red.

The two routes are leaked from the global route table into the Red VR and the connected subnet 8.0.0.0/24 is leaked from the Red VR to the global route table.

The following commands also add a non-leaked static route for the 56.6.6.0/24 subnetwork scoped to the domain of Red VR.

```
(Switch) (Config)#ip routing

(Switch) (Config)#interface 0/27

(Switch) (Interface 0/27)#routing

(Switch) (Interface 0/27)#ip vrf forwarding Red

Warning: routing interface moved from Default router instance to "Red" router in stance.

(Switch) (Interface 0/27)#ip address 8.0.0.1 /24

(Switch) (Interface 0/27)#interface 0/26

(Switch) (Interface 0/26)#routing

(Switch) (Interface 0/26)#ip address 9.0.0.1 /24

(Switch) (Interface 0/26)#exit

(Switch) (Config)#ip route 56.6.6.0 255.255.255.0 9.0.0.2
```

4. To leak routes from the global routing table to the VRF route table, use the following example:

```
(Switch) (Config)#ip route vrf Red 9.0.0.2 255.255.255.255 9.0.0.2 0/26

(Switch) (Config)#ip route vrf Red 56.6.6.0 255.255.255.0 9.0.0.2 0/26
```

To leak routes from the VRF's routing table to the global routing table, use the following example:

```
(Switch) (Config)#ip route 8.0.0.2 255.255.255.255 0/27
```

To leak routes (non-leaked) internal to the VRF's route table, use the following example:

```
(Switch) (Config)#ip route vrf Red 66.6.6.0 255.255.255.0 8.0.0.2
```

8. Configuring Multicast Routing

8.1. L3 Multicast Overview

IP Multicasting enables a network host (or multiple hosts) to send an IP datagram to multiple destinations simultaneously. The initiating host sends each multicast datagram only once to a destination multicast group address, and multicast routers forward the datagram only to hosts who are members of the multicast group. Multicast enables efficient use of network bandwidth because each multicast datagram needs to be transmitted only once on each network link, regardless of the number of destination hosts. Multicasting contrasts with IP unicasting, which sends a separate datagram to each recipient host. The IP routing protocols can route multicast traffic, but the IP multicast protocols handle the multicast traffic more efficiently with better use of network bandwidth.

Applications that often send multicast traffic include video or audio conferencing, Whiteboard tools, stock distribution tickers, and IP-based television (IP/TV).

8.1.1. IP Multicast Traffic

IP multicast traffic is traffic that is destined to a host group. Host groups are identified by class D IP addresses, which range from 224.0.0.0 to 239.255.255.255. When a packet with a broadcast or multicast destination IP address is received, the switch will forward a copy into each of the remaining network segments in accordance with the IEEE MAC Bridge standard. Eventually, the packet is made accessible to all nodes connected to the network.

This approach works well for broadcast packets that are intended to be seen or processed by all connected nodes. In the case of multicast packets, however, this approach could lead to less efficient use of network bandwidth, particularly when the packet is intended for only a small number of nodes. Packets will be flooded into network segments where no node has any interest in receiving the packet. The L3 multicast features on the switch help to ensure that only the hosts in the multicast group receive the multicast traffic for that group.

Multicast applications send one copy of a packet, and address it to a group of receivers (Multicast Group Address) rather than to a single receiver (unicast address). Multicast depends on the network to forward the packets to only those networks and hosts that need to receive them.

8.1.2. Multicast Protocol Switch Support

Multicast protocols are used to deliver Multicast packets from one source to multiple receivers. The following table summarizes the multicast protocols that the switch supports.

Protocol	IPv4 or IPv6	For Communication Between
IGMP	IPv4	Host-to-L3 switch/router
MLD	IPv6	Host-to-L3 switch (router)
PIM-SM	IPv4 or IPv6	L3-switch/router-to-L3 switch/router

Table 8-1: Multicast Protocol Support Summary

8.1.3. Multicast Protocol Roles

Hosts must have a way to identify their interest in joining any particular multicast group, and routers must have a way to collect and maintain group memberships. These functions are handled by the IGMP protocol in IPv4. In IPv6, multicast routers use the Multicast Listener Discover (MLD) protocol to maintain group membership information.

Multicast routers must also be able to construct a multicast distribution tree that enables forwarding multicast datagrams only on the links that are required to reach a destination group member. Protocols such as PIM handles this function.

IGMP and MLD are multicast group discovery protocols that are used between the clients and the local multicast router. PIM-SM, and PIM-DM are multicast routing protocols that are used across different subnets, usually between the local multicast router and remote multicast router.

8.1.4. Multicast Switch Requirements

You use the IPv4/IPv6 multicast feature on the switch to route multicast traffic between VLANs on the switch. If all hosts connected to the switch are on the same subnet, there is no need to configure the IP/IPv6 multicast feature. If the switch does not handle L3 routing, you can use IGMP snooping or MLD snooping to manage port-based multicast group membership. For more information, see “IGMP Snooping”. If the local network does not have a multicast router, you can configure the switch to act as the IGMP querier. For more information, see “IGMP Snooping Querier”.

If the switch is configured as a L3 switch and handles inter-VLAN routing through static routes or OSPF and multicast traffic is transmitted within the network, enabling and configuring L3 multicast routing on the switch is recommended.

8.1.5. Determining which Multicast Protocols to Enable

IGMP is recommended on any switch that participates in IPv4 multicasting. MLD is recommended on any switch that participates in IPv6 multicasting. PIM-DM and PIM-SM are multicast routing protocols that help determine the best route for IP (PIM) and IPv6 (PIM) multicast traffic.

8.1.6. Multicast Routing Tables

Multicast capable/enabled routers forward multicast packets based on the routes in the Multicast Routing Information Base (MRIB). These routes are created in the MRIB during the process of building multicast distribution trees by the Multicast Protocols running on the router. Different IP Multicast routing protocols use different techniques to construct these multicast distribution trees.

8.1.7. Multicast Tunneling

If Multicast traffic is to be routed through a part of a network that does not support multicasting (routers which are not multicast capable) then the multicast packets are encapsulated in an IP datagram and sent as a unicast packet. When the multicast router at the remote end of the tunnel receives the packet, the router

strips off the IP encapsulation and forwards the packet as an IP Multicast packet. This process of encapsulating multicast packets in IP is called tunneling.

8.1.8. IGMP

The Internet Group Management Protocol (IGMP) is used by IPv4 systems (hosts, L3 switches, and routers) to report their IP multicast group memberships to any neighboring multicast routers. The switch performs the multicast router role of the IGMP protocol, which means it collects the membership information needed by the active multicast routing protocol.

The switch supports IGMP Version 3. Version 3 adds support for source filtering, which is the ability for a system to report interest in receiving packets only from specific source addresses, as required to support Source-Specific Multicast [SSM], or from all but specific source addresses, sent to a particular multicast address. Version 3 is designed to be interoperable with Versions 1 and 2.

8.1.8.1. IGMP Proxy

IGMP proxy enables a multicast router to learn multicast group membership information and forward multicast packets based upon the group membership information. The IGMP Proxy is capable of functioning only in certain topologies that do not require Multicast Routing Protocols (i.e., PIM-DM and PIM-SM) and have a tree-like topology, as there is no support for features like reverse path forwarding (RPF) to correct packet route loops.

The proxy contains many downstream interfaces and a unique upstream interface explicitly configured. It performs the host side of the IGMP protocol on its upstream interface and the router side of the IGMP protocol on its downstream interfaces.

The IGMP proxy offers a mechanism for multicast forwarding based only on IGMP membership information. The router must decide about forwarding packets on each of its interfaces based on the IGMP membership information. The proxy creates the forwarding entries based on the membership information and adds it to the multicast forwarding cache (MFC) in order not to make the forwarding decision for subsequent multicast packets with same combination of source and group.

8.1.9. MLD Protocol

MLD, Multicast Listener Discovery is a protocol for IPv6 multicast router. It is to discover the presence of multicast listener on the local network and it is also used to discover which multicast packets are of interest to neighboring nodes. MLD is a sub-protocol of ICMP v6, Internet Control Message Protocol version 6. MLD messages are a subset of ICMP v6 messages.

Router use MLD to learn whether multicast group members of a group are present in their directly connected network. Each host in the multicast network sends MLD join report to join the multicast group and send MLD Leave message at any time as IGMP protocol act. The Multicast packets are delivered to a group using IPv6 unicast packet.

MLD v1 (RFC 2710) is equivalent to IGMP v2. MLDv2 (RFC 3810) is equivalent to IGMP v3 (MLDv2 is fully backward compatible with MLD version 1). All MLD messages are link-local with a hop limit of 1, and they all have “router alter option” set.

Router Alter Option implies a hop by hop option header. MLD 3 types of messages:

- General Query: multicast address field is set to 0 (::), is for learning which multicast addresses have listeners on the subnet.
- Group-Specific Query:
- Group-and-Source-Specific Query

8.1.9.1. Join Mechanism

When receiving an IGMP query message, clients will respond with IGMP Join Report for the group it is interested. MLD reports must be sent with a valid IPv6 link-local source address, or, if the outgoing interface has not yet acquired a valid link-local address, the unspecific address(::). Sending reports with the unspecified address is allowed to support the use of IPv6 multicast in the Neighbor Discovery Protocol

8.1.9.2. Leave Mechanism

Multicast client can leave multicast group any time by sending MLD Done message or sending the MLD listener report that excludes the group to the link-scope all routers multicast address FF02::2.

If multicast clients leave multicast group quietly without sending notification to the multicast router, the multicast router stops forwarding traffic after client response timeout.

8.1.10. PIM Protocol

The Protocol Independent Multicast protocol is a simple, protocol-independent multicast routing protocol. PIM uses an existing unicast routing table and a Join/Prune/Graft mechanism to build a tree. PIM switches support two types of PIM: sparse mode (PIM-SM) and dense mode (PIM-DM).

PIM-SM is most effective in networks with a sparse population of multicast receivers. In contrast, PIM-DM is most effective in networks with densely populated multicast receivers. In other words, PIM-DM can be used if the majority of network hosts request to receive a multicast stream, while PIM-SM might be a better choice in networks in which a small percentage of network hosts, located throughout the network, wish to receive the multicast stream.

Note: The switch supports PIM-SM only.

8.1.10.1. Using PIM-SM as the Multicast Routing Protocol

PIM-SM is used to efficiently route multicast traffic to multicast groups that may span wide area networks where bandwidth is a constraint.

PIM-SM uses shared trees by default and implements source-based trees for efficiency; it assumes that no hosts want the multicast traffic unless they specifically ask for it. It creates a shared distribution tree centered on a defined rendezvous point (RP) from which source traffic is relayed to the receivers. Senders first send the multicast data to the RP, which in turn sends the data down the shared tree to the receivers.

Shared trees centered on an RP do not necessarily provide the shortest, most optimal path. In such cases, PIM-SM provides a means to switch to more efficient source-specific trees. A data threshold rate is configured to determine when to switch from shared-tree to source-tree.

PIM-SM uses a Bootstrap Router (BSR), which advertises information to other multicast routers about the RP. In a given network, a set of routers can be administratively enabled as candidate bootstrap routers. If it is not apparent which router should be the BSR, the candidates flood the domain with advertisements. The router with the highest priority is elected. If all the priorities are equal, then the candidate with the highest IP address becomes the BSR.

Only one RP address can be used at a time within a PIM domain. You can configure a static RP on the switch. However, if the PIM domain uses the BSR to dynamically learn the RP, configuring a static RP is not required. By default the RP advertised by the BSR is used, but you can specify that the static RP to override any dynamically learned RP from the BSR.

If an interface on a switch configured with PIM-SM neighbors another PIM-SM domain, the PIM BSR messages should not flood into the neighboring PIM domain because the neighbor domain might not share the same set of RPs, candidate RPs, BSR, and candidate BSRs. The switch software allows you to configure an interface that borders the PIM boundary prevent transmission (sending and receiving) of PIM BSR messages. PIM-SM is defined in RFC 4601.

8.2. Default L3 Multicast Values

IP and IPv6 multicast is disabled by default. The following table shows the default values for L3 multicast and the multicast protocols.

Parameter	*Default Value*
IPv4 Multicast Defaults	
L3 Multicast Admin Mode	Disabled
Maximum Multicast Routing Table Entries	2048
Static Multicast Routes	None configured
Interface TTL Threshold	1
IGMP Defaults	
IGMP Admin Mode	Disabled globally and on all interfaces
IGMP Version	v3
IGMP Robustness	2
IGMP Query Interval	125 seconds
IGMP Query Max Response Time	10 seconds
IGMP Startup Query Interval	31 seconds
IGMP Startup Query Count	2
IGMP Last Member Query Interval	1 second
IGMP Last Member Query Count	2
IGMP Proxy Interface Mode	Disabled
IGMP Proxy Unsolicited Report Interval	1 second
MLD Defaults	
MLD Admin Mode	Disabled globally and on all interfaces
MLD Version	v2
MLD Query Interval	125 seconds
MLD Query Max Response Time	10,000 milliseconds
MLD Last Member Query Interval	1000 milliseconds
MLD Last Member Query Count	2
MLD Proxy Interface Mode	Disabled
MLD Proxy Unsolicited Report Interval	1 second
PIM Defaults	
PIM Protocol	Disabled globally and on all interfaces
PIM-SM Data Threshold Rate	0 Kpbs
PIM-SM Register Threshold Rate	0 Kbps
PIM Hello Interval	30 seconds (when enabled on an interface)
PIM-SM Join/Prune Interval	60 seconds (when enabled on an interface)
PIM-SM BSR Border	Disabled
PIM-SM DR Priority	1 (when enabled on an interface)
PIM Candidate Rendezvous Points (RPs)	None configured
PIM Static RP	None configured
PIM Source-Specific Multicast (SSM) Range	None configured. Default SSM group address is 232.0.0.0/8 for IPv4 multicast and ff3x::/32 for IPv6 multicast.
PIM BSR Candidate Hash Mask Length	30 (IPv4) 126 (IPv6)
PIM BSR Candidate Priority	0
DVMRP Defaults	
DVMRP Admin Mode	Disabled globally and on all interfaces
DVMRP Version	3
DVMRP Interface Metric	1

Table 8-2: L3 Multicast Defaults

8.3. L3 Multicast Configuration Examples

8.3.1. Configuring Multicast VLAN Routing with IGMP and PIM-SM

This example describes how to configure a switch with two VLAN routing interfaces that route IP multicast traffic between the VLANs. PIM and IGMP are enabled on the switch and interfaces to manage the multicast routing. IGMP snooping is enabled on the VLAN interfaces to control the multicast subscriptions within each VLAN.

VLAN 10 is statically configured as the RP for the multicast group.

Note: PIM does not require OSPF specifically; static routing could also be configured for unicast routing.

The configuration in this example takes place on L3 Switch A shown in the following figure. The red arrows indicate the path that multicast traffic takes. L3 Switch A is configured as the RP for the PIM domain, so it is in charge of sending the multicast stream to L3 Switch B and L3 Switch C, and these switches forward the multicast data to the hosts that have requested to receive the data.

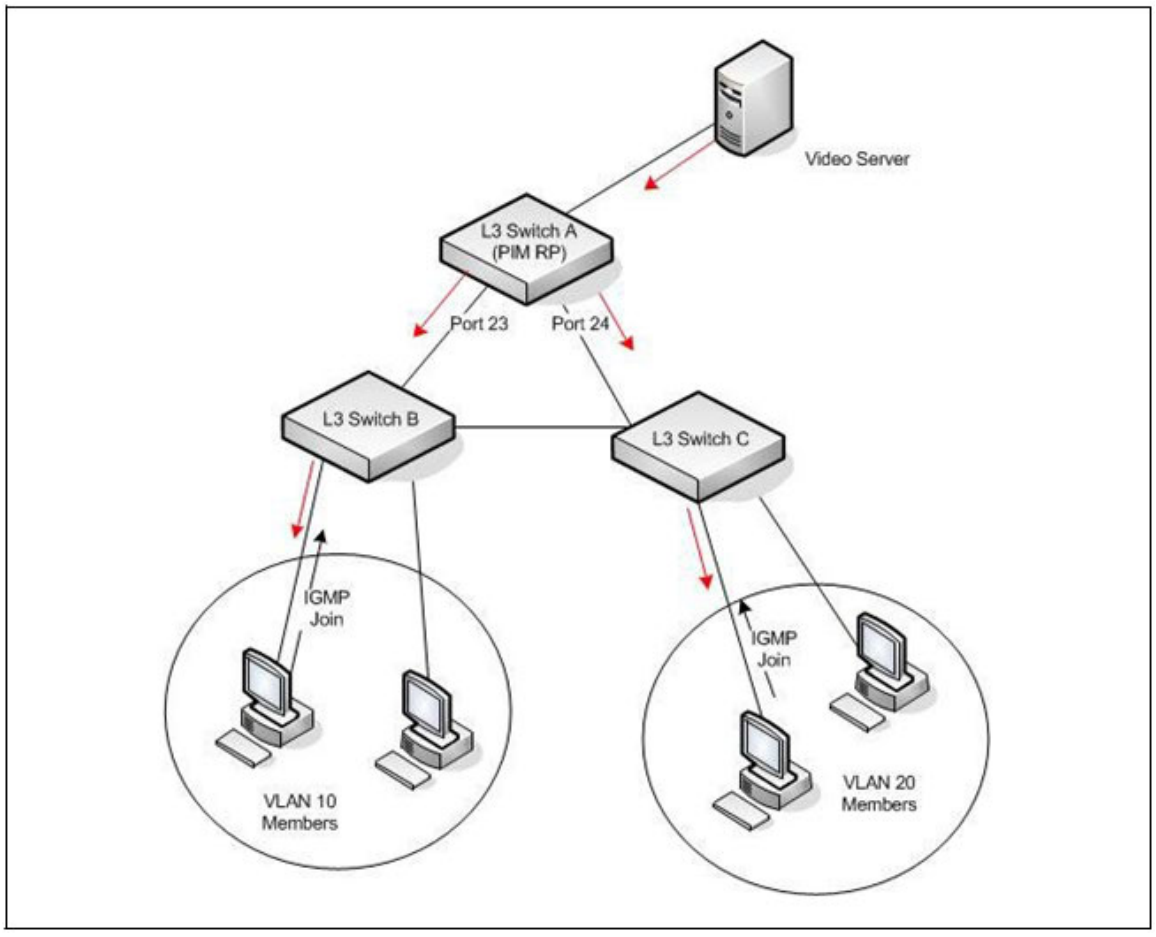


Figure 8-1: Multicast VLAN Routing with IGMP and PIM-SM Example

In addition to multicast configuration, this example includes commands to configure STP and OSPF on L3

Switch A. STP is configured on the ports that connect the switch to other switches. OSPF is configured to route unicast traffic between the VLANs.

To configure the switch:

1. Create two VLANs and configure them as routing VLANs.

```
(Switch) (Config)#vlan database
(Switch) (Vlan)#vlan 10,20
(Switch) (Vlan)#exit
(Switch) (Config)#interface vlan 10
(Switch) (Config)#interface vlan 20
```

2. While in VLAN Database mode, enable IGMP snooping on the VLANs.

```
(Switch) (Vlan)#set igmp 10
(Switch) (Vlan)#set igmp 20
(Switch) (Vlan)#exit
```

3. Add VLANs to interfaces 0/23 and 0/24.

```
(Switch) (Config)#interface 0/23
(Switch) (Interface 0/23)#switchport allowed vlan add 10
(Switch) (Interface 0/23)#switchport native vlan 10
(Switch) (Interface 0/23)#switchport allowed vlan remove 1
(Switch) (Interface 0/23)#exit
(Switch) (Config)#interface 0/24
(Switch) (Interface 0/24)#switchport allowed vlan add 20
(Switch) (Interface 0/24)#switchport native vlan 20
(Switch) (Interface 0/24)#switchport allowed vlan remove 1
(Switch) (Interface 0/24)#exit
```

4. Enable routing on the switch and configure the OSPF router ID.

```
(Switch) (Config)#ip routing
(Switch) (Config)#router ospf
(Switch) (Config-router)#router-id 3.3.1.1
(Switch) (Config-router)#exit
```

5. Configure VLAN 10 as a VLAN routing interface and specify the OSPF area. When you assign an IP address to the VLAN, routing is automatically enabled.

```
(Switch) (Config)#interface vlan 10
(Switch) (if-vlan10)#ip address 192.168.10.4 255.255.255.0
(Switch) (if-vlan10)#ip ospf area 0
```

6. Enable IGMPv2 and PIM-SM on the VLAN routing interface.

```
(Switch) (if-vlan10)#ip igmp
(Switch) (if-vlan10)#ip igmp version 2
(Switch) (if-vlan10)#ip pim
(Switch) (if-vlan10)#exit
```

7. Configure VLAN 20 as a VLAN routing interface and specify the OSPF area.

```
(Switch) (Config)#interface vlan 20
(Switch) (if-vlan20)#ip address 192.168.20.4 255.255.255.0
(Switch) (if-vlan20)#ip ospf area 0
```

8. Enable IGMPv2 and PIM-SM on the VLAN routing interface.

```
(Switch) (if-vlan20)#ip igmp
(Switch) (if-vlan20)#ip igmp version 2
(Switch) (if-vlan20)#ip pim
(Switch) (if-vlan20)#exit
```

9. Globally enable IGMP snooping, IP multicast, IGMP, and PIM-SM on the switch.

```
(Switch) (Config)#ip igmp snooping
(Switch) (Config)#ip multicast
(Switch) (Config)#ip igmp
(Switch) (Config)#ip pim sparse
```

10. Configure VLAN 10 as the RP and specify the range of multicast groups for PIM-SM to control.

```
(Switch) (Config)#ip pim rp-address 192.168.10.4 225.0.0.0 240.0.0.0
```

8.3.2. Example 1: MLDv1 Configuration

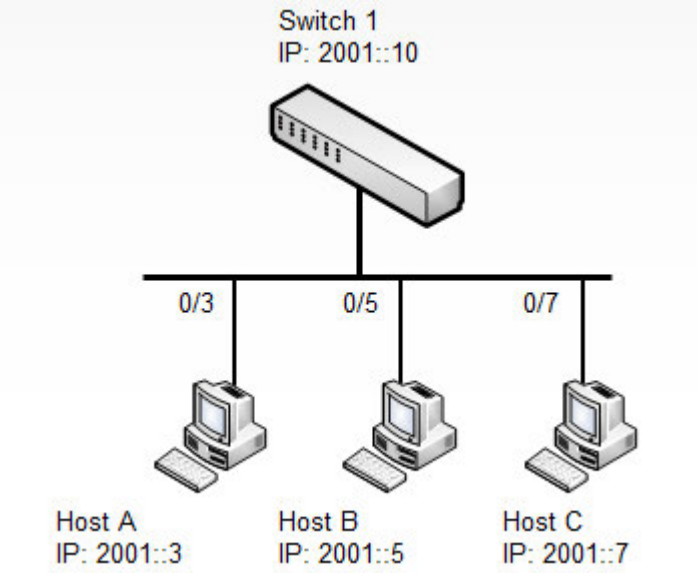


Figure 8-2: MLD Topology

To configure the switch:

Switch-1 Configuration

Step 1. Enable MLD and relative routing command on global mode

```
(Switch-1) (Config)#ip routing
(Switch-1) (Config)#ipv6 mld router
(Switch-1) (Config)#ipv6 unicast-routing
(Switch-1) (Config)#ip multicast
```

Step 2. Enable MLD on specific interface or VLAN interface

```
(Switch-1) (Config)#interface range 0/3-0/9
(Switch-1) (Interface 0/3-0/9)#ipv6 mld router
(Switch-1) (Interface 0/3-0/9)#ipv6 mld version 1
```

OR

```
(Switch-2) (Config)#interface vlan 1
(Switch-2) (if-vlan1)#ipv6 mld router
(Switch-2) (if-vlan1)#ipv6 mld version 1
```

8.3.3. Example 2: MLDv2 Configuration

Switch-1 Configuration

Step 1. Enable MLD and relative routing command on global mode

```
(Switch-1) (Config)#ip routing
(Switch-1) (Config)#ipv6 mld router
(Switch-1) (Config)#ipv6 unicast-routing
(Switch-1) (Config)#ip multicast
```

Step 2. Enable MLD on specific interface or VLAN interface

```
(Switch-1) (Config)#interface range 0/3-0/9
(Switch-1) (Interface 0/3-0/9)#ipv6 mld router
(Switch-1) (Interface 0/3-0/9)#ipv6 mld version 2
```

or

```
(Switch-2) (Config)#interface vlan 1
(Switch-2) (if-vlan1)#ipv6 mld router
(Switch-2) (if-vlan1)#ipv6 mld version 2
```

8.3.4. Example 3: MLD Configuration Verification

Verify IPv6 MLD VLAN interface configuration

```
(Switch-1) (Config)#show ipv6 mld interface vlan 1
Interface   .................................vlan 1
MLD Global Admin Mode. .......................Enabled
MLD Interface Admin Mode  ....................Enabled
MLD Operational Mode..........................Enabled
MLD Version...................................1
Query Interval (secs)   ......................125
Query Max Response Time(milli-secs) ..........10000
Robustness....................................2
Startup Query Interval (secs)  ...............31
Startup Query Count ..........................2
Last Member Query Interval (milli-secs) ........1000
Last Member Query Count   .....................2
```

Check multicast group information under vlan 1

```
(Switch-1) (Config)#show ipv6 mld groups vlan 1
Group Address.............................. ff15::777
Interface.................................. vlan 1
Up Time (hh:mm:ss)......................... 00:00:19
Expiry Time (hh:mm:ss)..................... 00:04:03
```

Check multicast group “ff15::777” information

```
(Switch1) (Config)#show ipv6 mld groups ff15::777
Interface.................................. vlan 1
Group Address.............................. ff15::777
Last Reporter.............................. fe80::1
Up Time (hh:mm:ss)......................... 00:00:34
Expiry Time (hh:mm:ss)..................... 00:03:48
Filter Mode................................ ------
Version1 Host Timer........................ 00:03:48
Group compat mode.......................... v1
```

9. Configuring Data Center Features

9.1. Data Center Technology Overview

The switch supports Data Center Bridging (DCB) features to increase the reliability of Ethernet-based networks in the data center. The Ethernet enhancements that DCB provides are well suited for Fiber Channel over Ethernet (FCoE) environments.

The following table provides a summary of the features this section describes.

Address	Description
PFC	If congestion occurs, Priority-Based Flow Control (PFC) provides a way to distinguish which traffic on a .physical link is paused, based on the priority of the traffic. See “Priority-Based Flow Control.”
DCBX	The Data Center Bridging Exchange (DCBX) protocol allows Data Center Bridging (DCB) devices to exchange configuration information with directly-connected peers, using type length value (TLV) information elements over LLDP. See “Data Center Bridging Exchange Protocol.”
CoS Queuing	Lets you directly configure some device queuing features to establish QoS behavior for different types of network traffic. See “CoS Queuing.”
ETS	Supports the Enhanced Transmission Selection (ETS) configuration and application priority TLVs that come from auto-upstream devices and are propagated to auto-downstream devices. See “Enhanced Transmission Selection.”
VXLAN Gateway	Enables VXLAN network virtualization network technologies to communicate with other network components (in particular, VLANs) and supports VTEP functionality for VXLAN tunnels on the switch. See “VXLAN Gateway Operation and Configuration.”

Table 9-1: DCB Features

9.2. Priority-based Flow Control

Ordinarily, when flow control is enabled on a physical link, it applies to all traffic on the link. When congestion occurs, the hardware sends pause frames that temporarily suspend traffic flow to help prevent buffer overflow and dropped frames.

PFC provides a means of pausing individual priorities within a single physical link. By pausing the congested priority or priorities independently, protocols that are highly loss-sensitive can share the same link with traffic that has different loss tolerances.

This feature is used in networks where the traffic has differing loss tolerances. For example, Fiber Channel traffic is highly sensitive to traffic loss. If a link contains both loss-sensitive data and other less loss-sensitive data, the loss-sensitive data should use a no-drop priority that is enabled for flow control.

Priorities are differentiated by the priority field of the IEEE 802.1Q VLAN header, which identifies an IEEE 802.1p priority value. These priority values must be mapped to internal class-of-service (CoS) values.

The PFC feature allows you to specify the CoS values that should be paused (due to greater loss sensitivity) instead of dropped when congestion occurs on a link. Unless configured as no-drop, all CoS priorities are considered non-pausable ("drop") when priority-based flow control is enabled until no-drop is specifically turned on.

9.2.1. PFC Operation and Behavior

PFC uses a new control packet defined in IEEE 802.1Qbb and therefore is not compatible with IEEE 802.3

Annex 31B flow control. An interface that is configured for PFC will be automatically disabled for flow control. When PFC is disabled on an interface, the flow control configuration for the interface becomes active. Any flow control frames received on a PFC configured interface are ignored.

Each priority is configured as either *drop* or *no-drop*. If a priority that is designated as no-drop is congested, the priority is paused. Drop priorities do not participate in pause. You must configure the same no-drop priorities across the network in order to ensure end-to-end lossless behavior.

Operator configuration of PFC is used only when the port is configured in a manual role. When interoperating with other equipment in a manual role, the peer equipment must be configured with identical PFC priorities and VLAN assignments. Interfaces not enabled for PFC ignore received PFC frames. Ports configured in auto-upstream or auto-downstream roles receive their PFC configuration from the configuration source and ignore any manually-configured information.

Note: This feature is configurable on physical full duplex interfaces only. To enable PFC on a Port-channel interface, the member interfaces must have the same configuration.

When PFC is disabled, the interface defaults to the IEEE 802.3 flow control setting for the interface. PFC is disabled by default.

If you enable priority-based flow control for a particular priority value on an interface, make sure 802.1p priority values are mapped to CoS values (see "CoS").

9.2.2. Configuring PFC

The network in this example handles standard data traffic and traffic that is time sensitive (such as voice and video). The time-sensitive traffic requires a higher priority than standard data traffic. All time-sensitive traffic is configured to use VLAN 100 and has an 802.1p priority of 5, which is mapped to hardware queue 4. The hosts that frequently send and receive the time-sensitive traffic are connected to ports 3, 5, and 10, so PFC is enabled on these ports with 802.1p priority 5 traffic as no-drop. The configuration also enables VLAN tagging so that the 802.1p priority is identified. This example assumes that VLAN 100 has already been configured.

Caution! All ports may be briefly shutdown when modifying either flow control or PFC settings. PFC uses a control packet defined in 802.1Qbb and is not compatible with 802.3x FC.

1. Map 802.1p priority 5 to traffic class 4. For more information about traffic classes, see "CoS".

```
(Switch) #configure
(Switch) (Config)#queue cos-map all 5 4
```

2. Enter Interface Configuration mode for ports 3, 5, and 10.

```
(Switch) (Config)#interface range 0/3,0/5,0/10
```

3. Enable PFC and configure traffic marked with 802.1p priority 5 to be paused rather than dropped when congestion occurs.

```
(Switch) (Interface 0/3,0/5,0/10)#data-center-bridging
(Switch) (Config-if-dcb)#priority-flow-control mode on
(Switch) (Config-if-dcb)#priority-flow-control priority 5 no-drop
(Switch) (config-if-dcb)#exit
```

4. Enable VLAN tagging on the ports so the 802.1p priority is identified.

```
(Switch) (Interface 0/3,0/5,0/10)#switchport allowed vlan add tagged 100
(Switch) (Interface 0/3,0/5,0/10)#exit
```

9.3. Data Center Bridging Exchange Protocol

The Data Center Bridging Exchange Protocol (DCBX) is used by DCB devices to exchange configuration information with directly connected peers. DCBX uses type-length-value (TLV) information elements over LLDP to exchange information, so LLDP must be enabled on the port to enable the information exchange. By default, LLDP is enabled on all ports. For more information, see “LLDP and LLDP-MED”.

The main objective of DCBX is to perform the following operations:

- **Discovery of DCB capability in a peer**: DCBX is used to learn about the capabilities of the peer device. It is a means to determine if the peer device supports a particular feature such as PFC.
- **DCB feature misconfiguration detection**: DCBX can be used to detect misconfiguration of a feature between the peers on a link. Misconfiguration detection is feature-specific because some features may allow asymmetric configuration.
- **Peer configuration of DCB features**: DCBX can be used by a device to perform configuration of DCB features in its peer device if the peer device is willing to accept configuration.

DCBX is expected to be deployed in Fiber Channel over Ethernet (FCoE) topologies in support of lossless operation for FCoE traffic. In these scenarios, all network elements are DCBX enabled. In other words, DCBX is enabled end-to-end.

The DCBX protocol supports the propagation of configuration information for the following features:

- Enhanced Transmission Selection (ETS)
- Priority-based Flow Control (PFC)
- Application Priorities

These features use DCBX to send and receive device configuration and capability information to the peer DCBX device.

The Application Priorities information is simply captured from the peer and potentially propagated to other peers by the DCBX component.

9.3.1. Interoperability with IEEE DCBX

To be interoperable with legacy industry implementations of DCBX protocol, the switch uses a hybrid model to support both the IEEE version of DCBX (IEEE 802.1Qaz) and legacy DCBX versions.

The switch automatically detects if a peer is operating with either of the two CEE DCBX versions or the IEEE standard DCBX version. This is the default mode. You can also configure DCBX to manually select one of the legacy versions or IEEE standard mode. In auto-detect mode, the switch starts operating in IEEE DCBX mode on a port, and if it detects a legacy DCBX device based on the OUI of the organization TLV, then the switch changes its DCBX mode on that port to support the version detected. There is no timeout mechanism to move back to IEEE mode. Once the DCBX peer times out, multiple peers are detected, the link is reset (link down/up) or as commanded by the operator, DCBX resets its operational mode to IEEE.

The interaction between the DCBX component and other components remains the same irrespective of the operational mode it is executing. For instance DCBX component interacts with PFC to get needed information to pack the TLVs to be sent out on the interface. Based on the operational control mode of the port, DCBX packs it in the proper frame format.

9.3.2. DCBX and Port Roles

Each port's behavior is dependent on the operational mode of that port and of other ports in the switch. The port mode is a DCBX configuration item that is passed to the DCBX clients to control the processing of their configuration information. There are four port roles:

- Manual
- Auto-Upstream
- Auto-Downstream
- Configuration Source

Ports operating in the manual role do not have their configuration affected by peer devices or by internal propagation of configuration. These ports have their operational mode, traffic classes, and bandwidth information specified explicitly by the operator. These ports advertise their configuration to their peer if DCBX is enabled on that port. Incompatible peer configurations are logged and counted with an error counter.

The default operating mode for each port is manual. A port that is set to manual mode sets the willing bit for DCBX client TLVs to false. Manually- configured ports never internally propagate or accept internal or external configuration from other ports, in other words, a manual configuration discards any automatic configuration. Manually-configured ports may notify the operator of incompatible configurations if client configuration exchange over DCBX is enabled. Manually- configured ports are always operationally enabled for DCBX clients,

regardless of whether DCBX is enabled. Operationally enabled means that the port reports that it is able to operate using the current configuration.

A port operating in the auto-upstream role advertises a configuration, but it is also willing to accept a configuration from the link-partner and propagate it internally to the auto-downstream ports as well as receive configuration propagated internally by other auto-upstream ports. Specifically, the willing parameter is enabled on the port and the recommendation TLV is sent to the peer and processed if received locally. The first auto- upstream port to successfully accept a compatible configuration becomes the configuration source. The configuration source propagates its configuration to other auto-upstream and auto-downstream ports. Only the configuration source may propagate configuration to other ports internally. Auto-upstream ports that receive internally propagated information ignore their local configuration and utilize the internally propagated information.

Peer configurations received on auto-upstream ports other than the configuration source result in one of two possibilities. If the configuration is compatible with the configuration source, then the DCBX client becomes operationally active on the upstream port. If the configuration is not compatible with the configuration source, then a message is logged indicating an incompatible configuration, an error counter is incremented, and the DCBX client is operationally disabled on the port. The expectation is that the network administrator configures the upstream devices appropriately so that all such devices advertise a compatible configuration.

A port operating in the auto-downstream role advertises a configuration but is not willing to accept one from the link partner. However, the port will accept a configuration propagated internally by the configuration source. Specifically, the willing parameter is disabled on auto-downstream. By default, auto-downstream ports have the recommendation TLV parameter enabled. Auto-downstream ports that receive internally propagated information ignore their local configuration and utilize the internally propagated information. Auto-downstream ports propagate PFC, ETS, and application priority information received from the configuration source.

In the Configuration Source role, the port has been manually selected to be the configuration source. Configuration received over this port is propagated to the other auto-configuration ports, however, no automatic election of a new configuration source port is allowed. Events that cause selection of a new configuration source are ignored. The configuration received over the configuration source port is maintained until cleared by the operator (set the port to the manual role).

9.3.3. Configuration Source Port Selection Process

When an auto-upstream or auto-downstream port receives a configuration from a peer, the DCBX client first checks if there is an active configuration source. If there is a configuration source already selected, the received configuration is checked against the local port operational values as received from the configuration source, and if compatible, the client marks the port as operationally enabled. If the configuration received from the peer is determined to not be compatible, a message is logged, an error counter is incremented and the DCBX clients become operationally disabled on the port. Operationally disabled means that PFC will not operate over the port. The port continues to keep link up and exchanges DCBX packets. If a compatible configuration is later received, the DCBX clients will become operationally enabled.

If there is no configuration source, a port may elect itself as the configuration source on a first-come, first-serve basis from the set of eligible ports. A port is eligible to become the configuration source if the following conditions are true:

- No other port is the configuration source.
- The port role is auto-upstream.
- The port is enabled with link up and DCBX enabled.
- The port has negotiated a DCBX relationship with the partner.
- The switch is capable of supporting the received configuration values, either directly or by translating the values into an equivalent configuration.

Whether or not the peer configuration is compatible with the configured values is NOT considered.

The newly elected configuration source propagates DCBX client information to the other ports and is internally marked as being the port over which configuration has been received. Configuration changes received from the peer over the configuration source port are propagated to the other auto-configuration ports. Ports receiving auto-configuration information from the configuration source ignore their current settings and utilize the configuration source information.

When a configuration source is selected, all auto-upstream ports other than the configuration source are marked as willing disabled.

To reduce flapping of configuration information, if the configuration source port is disabled, disconnected or loses LLDP connectivity, the system clears the selection of configuration source port (if not manually selected) and enables the willing bit on all auto-upstream ports. The configuration on the auto-configuration ports is not cleared (configuration holdover). If the user wishes to clear the configuration on the system in this scenario, the user can put the configuration source port into manual mode.

When a new port is selected as configuration source, it is marked as the configuration source, the DCBX configuration is refreshed on all auto-configuration ports and each port may begin configuration negotiation with their peer again (if any information has changed).

9.3.4. Configuring DCBX

In this example, port 0/1 on the M4500 series switch connects to an FCoE-facing (FCF) switch. This port is designated as default DCBX auto-upstream ports. Port 0/2 on the M4500 series switch is directly connected to a Converged Network Adapter (CNA) on a network server. The configuration advertised by the FCF is distributed from port

0/1 to port 0/2. In order to reduce configuration flapping, ports that obtain configuration information from a configuration source port will maintain that configuration for 2× the LLDP timeout, even if the configuration source port becomes operationally disabled.

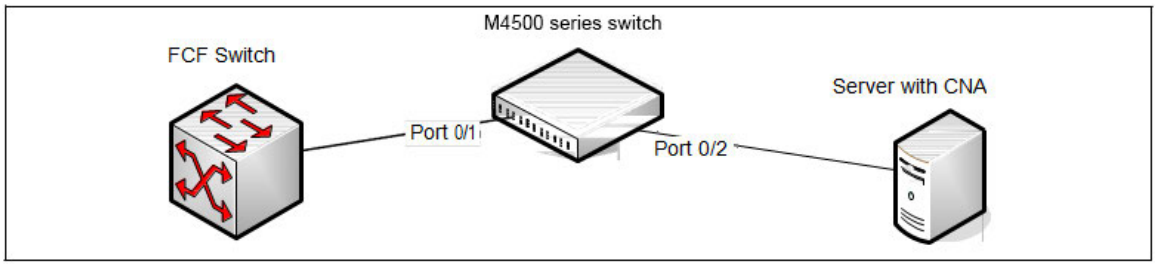


Figure 9-1: DCBX Configuration

1. Map 802.1p priority 3 to traffic class 3. For more information about traffic classes, see “CoS”.

```
(Switch) #configure
(Switch) (Config)#queue cos-map all 3 3
```

2. Enter Interface Configuration mode for port 1.

```
(Switch) (Config)#interface 0/1
```

3. Enable the LLDP transmit and receive capability on the port.

```
 (Switch) (Interface 0/1)#lldp transmit
(Switch) (Interface 0/1)#lldp receive
```

4. Enable the port as the configuration source. This port is connected to a trusted FCF. Configuration received over this port is propagated to the other auto-configuration ports.

```
(Switch) (Interface 0/1)#lldp dcbx port-role configuration-source
(Switch) (Interface 0/1)#exit
```

5. Enter Interface Configuration mode for port 2.

```
(Switch) (Config)#interface 0/2
```

6. Enable the LLDP transmit and receive capability on the port.

```
(Switch) (Interface 0/2)#lldp transmit
(Switch) (Interface 0/2)#lldp receive
```

7. Configure the LLDP port role as *auto-down*, which means the port advertises a configuration but is not willing to accept one from the link partner. However, the port will accept a configuration propagated internally by the configuration source (port 0/1).

```
(Switch) (Interface 0/2)#lldp dcbx port-role auto-down
(Switch) (Interface 0/2)#exit
```

9.4. CoS Queuing

In a typical switch or router, each physical port consists of one or more queues for transmitting packets on the attached network. Multiple queues per port are often provided to give preference to certain packets over others based on user-defined criteria. When a packet is queued for transmission in a port, the rate at which it is serviced depends on how the queue is configured—and possibly the amount of traffic present in the other

queues of the port. If a delay is necessary, packets are held in the queue until the scheduler authorizes the queue for transmission. As queues become full, packets have no place to be held for transmission and get dropped by the device.

The drop precedence of a packet is an indication of whether the packet is more or less likely to be dropped during times of queue congestion. Often referred to as packet coloring, a low drop precedence (green) allows the packet to be transmitted under most circumstances, a higher drop precedence (yellow) subjects the packet to dropping when bursts become excessive, while the highest drop precedence (red) discards the packet whenever the queue is congested. In some hardware implementations, the queue depth can be managed using tail dropping or a weighted random early discard, or WRED, technique. These methods often use customizable threshold parameters that are specified on a per-drop-precedence basis.

The switch supports Differentiated Services (DiffServ), which allows traffic to be classified into streams and given certain QoS treatment in accordance with defined per-hop behaviors. However, the DiffServ feature does not offer direct configuration of the hardware COS queue resources.

The COS Queuing feature allows the switch administrator to directly configure certain aspects of device queuing to provide the desired QoS behavior for different types of network traffic when the complexities of DiffServ are not required. The priority of a packet arriving at an interface can be used to steer the packet to the appropriate outbound COS queue through a mapping table. With the CoS queuing feature, COS queue characteristics such as minimum guaranteed bandwidth, transmission rate shaping, etc. can be configured at the queue (or port) level.

With support for the multistage scheduling architecture, the COS queue feature provides a method to configure Traffic Class Groups (TCGs) to extend the COS queue management. Multiple COS queues can be mapped to a single TCG. Each TCG can have a configured minimum guaranteed bandwidth allocation and a scheduling algorithm similar to the COS queue configuration. The TCG scheduling and bandwidth enforcement occurs after the COS queue scheduling and bandwidth enforcement is performed. Therefore all COS queues mapped to the same TCG share the scheduling and bandwidth properties of the TCG.

9.4.1. CoS Queuing Function and Behavior

Like CoS mapping, CoS queuing uses the concept of trusted and untrusted ports. CoS queuing builds on includes user-configurable settings that affect hardware queue operation.

9.4.1.1. Trusted Port Queue Mappings

A trusted port is one that takes at face value a certain priority designation within arriving packets. Specifically, a port may be configured to trust one of the following packet fields:

- 802.1p User Priority
- IP Precedence
- IP DSCP

Packets arriving at the port ingress are inspected and their trusted field value is used to designate the COS queue that the packet is placed when forwarded to the appropriate egress port. A mapping table associates the trusted field value with the desired COS queue.

9.4.1.2. Un-trusted Port Default Priority

Alternatively, a port may be configured as un-trusted, whereby it does not trust any incoming packet priority designation and uses the port default priority value instead. All packets arriving at the ingress of an un-trusted port are directed to a specific COS queue on the appropriate egress port(s) in accordance with the configured default priority of the ingress port. This process is also used for cases where a trusted port mapping is unable to be honored, such as when a non-IP packet arrives at a port configured to trust the IP precedence or IP DSCP value.

9.4.1.3. Queue Configuration

Queue configuration involves setting the following hardware port egress queue configuration parameters:

- Scheduler type: strict vs. weighted
- Minimum guaranteed bandwidth
- Maximum allowed bandwidth (i.e. shaping)
- Queue management type: tail-drop vs. WRED
- Tail drop parameters: threshold
- WRED parameters: minimum threshold, maximum threshold, drop probability

Defining these settings on a per-queue basis allows the user to create the desired service characteristics for different types of traffic. The tail drop and WRED parameters are specified individually for each supported drop precedence level.

In addition, the following settings can be specified on a per-interface basis:

- Queue management type: tail drop vs. WRED (only if per-queue configuration is not supported)
- WRED decay exponent

9.4.1.4. Traffic Class Groups

For platforms such as M4500 series switches that support the multistage scheduling architecture, the Traffic Class Groups (TCGs) extend the egress queuing to make use of multiple levels of scheduling. A TCG defines a collection of egress COS Queues. The configuration parameters for the TCG specify the class of service characteristics applied to the aggregated traffic from the associated COS queues. This involves setting the following configuration parameters to each TCG.

- Map one or more COS queues to the TCG.
- Set the scheduling type for each TCG: Strict vs. WDRR
- Set the weight percentages for each TCG.

- Set the minimum guaranteed bandwidth for each TCG. The minimum bandwidth is specified in terms of the percentage of the total link bandwidth.

- Set the maximum allowed bandwidth for each TCG. The maximum bandwidth is specified in terms of the percentage of the total link bandwidth.

TCG configuration parameters are similar to that of COS queues. That is, the configuration of scheduling attributes such as minimum bandwidth, maximum bandwidth, and scheduling algorithm also apply to TCG. The behavior of a TCG with respect to scheduling algorithm and bandwidth allocation configuration is the same as that of COS Queues.

Each TCG is associated with a weight percentage which defines the priority of the TCG to be serviced when WDRR is configured as the scheduling type of the TCG. The weight of the TCG is used only after the minimum guaranteed bandwidth of each of the TCG is met and after all the strict priority TCGs are serviced. The weight of the TCG is then used to prioritize the TCGs among the TCGs that are configured for WDRR.

9.4.2. Configuring CoS Queuing and ETS

This example shows the manual configuration of the CoS queuing feature in a network where traffic needs to be prioritized based on the protocol frame-loss tolerance. For example, FCoE traffic is highly sensitive to traffic loss. If a port has both loss-sensitive data and other less loss-sensitive data, then the loss-sensitive data is categorized into the same TCG to provide control over the bandwidth allocation and scheduling for the loss-sensitive traffic.

In this example, loss-sensitive traffic is sent with an 801.p priority value of 4, and less loss-sensitive traffic is sent with an 801.p priority value priority of 1. The following steps show how to configure the switch to prioritize the traffic.

1. Configure one to one mapping between 802.1p priority and COS Queue on the ingress port. Frames with 802.1p priority 1 are assigned to COS 1 queue and similarly frames with 802.1p priority 2 are assigned to COS2 and so on.

```
(Switch) (Config)#queue cos-map all 0 0
(Switch) (Config)#queue cos-map all 1 1
(Switch) (Config)#queue cos-map all 2 2
(Switch) (Config)#queue cos-map all 3 3
(Switch) (Config)#queue cos-map all 4 4
(Switch) (Config)#queue cos-map all 5 5
(Switch) (Config)#queue cos-map all 6 6
(Switch) (Config)#queue cos-map all 7 7
```

2. Enable 802.1p Trust mode on all the ports.

```
(Switch) (Config)#interface range 0/1-0/16
(Switch) (Interface 0/1-0/16)#queue trust dot1p
(Switch) (Interface 0/1-0/16)#exit
```

3. Configure the mapping between COS queues and Traffic Classes Groups. Configure the Traffic Class Group that such 802.1p priority 4 is assigned to TCG1 and 802.1p priority 1 is assigned to TCG2 so that less loss sensitive traffic does not starve the loss sensitive traffic even during traffic bursts. Assign 802.1p priority 7 traffic to TCG0.

```
(Switch) (Config)#classofservice traffic-class-group 4 1
(Switch) (Config)#classofservice traffic-class-group 1 2
(Switch) (Config)#classofservice traffic-class-group 7 0
```

4. Enable VLAN tagging on the ports so the 802.1p priority is identified. The interfaces in this example are members of VLAN 100, which has been previously configured.

```
(Switch) (Config)#interface range 0/1-0/16
(Switch) (Interface 0/1-0/16)#switchport allowed vlan add tagged 100
(Switch) (Interface 0/1-0/16)#exit
```

5. Configure the weight percentage of TCG0 to 10%, and the weights of TCG1 and TCG2 to 45% each.

```
(Switch) (Config)#traffic-class-group weight 10 45 45
```

6. Associate weighted round robin scheduling with TCG1 and TCG2.

```
(Switch) (Config)#no traffic-class-group strict 1 2
```

7. Configure TCG0 for strict priority scheduling.

```
(Switch) (Config)#traffic-class-group strict 0
```

8. Associate TCG0 with CoS queue 7 so that it serves the high priority internal control traffic with CoS 7.

```
(Switch) (Config)#classofservice traffic-class-group 7 0
```

9. Configure the minimum bandwidth percentage for all the TCGs to be zero.

```
(Switch) (Config)#traffic-class-group min-bandwidth 0 0 0
```

After you perform all of the previous steps, the data traffic with an 802.1p priority is sent through TCG1, and 45% of the bandwidth (excluding TCG0 bandwidth) is reserved for TCG1. This protects the TCG1 traffic from traffic that is transmitted on TCG2. Any burst in traffic being transmitted in TCG2 does not affect traffic in TCG1. If TCG2 is not being utilized to the full potential then TCG1 can still use that bandwidth for transmitting TCG1 traffic.

With the configuration in this example, TCG0 with strict priority gets highest priority and can consume the full bandwidth of the pipeline. TCG1 and TCG2 share the remaining bandwidth after TCG0 consumes its share of the pipeline.

Based on this configuration, when the switch sends the configuration ETS TLVs to the peer, the values that are given to DCBX are as follows:

- **Willing Bit**—This bit is set to TRUE for auto-upstream interfaces if there is no configuration source or
- FALSE if there is a configuration source, and FALSE for auto-downstream and manual ports.
- **Credit-based Shaper support and Max TC**—These values are platform-specific.

- **Priority Assignment Table**—The following table contains the default values advertised by DCBX to the peer DCBX device. If available, the mapping translated from the configuration source is used. This table defines the mapping between the egress Traffic Class Group and ingress 802.1p priority.

802.1p Priority	Traffic Class
0	0
1	0
2	0
3	0
4	0
5	0
6	0
7	0

Table 9-2: 802.1p-to-TCG Mapping

- **TC Bandwidth And TSA Assignment Table**—The following table contains the default values advertised by DCBX to the peer DCBX device. If available, the assignments translated from the configuration source is used. This table defines the bandwidth allocated to each Traffic Class Group and the respective scheduling algorithm for each TCG; the scheduling algorithm is enumerated in the IEEE 802.1Q specification.

Traffic Class		Bandwidth percentage	Scheduling Algorithm
	0	10	strict priority (tail-drop) (0)
	1	45	strict priority (tail-drop) (0)
	2	45	strict priority (tail-drop) (0)

Table 9-3: TCG Bandwidth and Scheduling

9.5. Enhanced Transmission Selection

Enhanced Transmission Selection (ETS) enables the sharing and redistribution of network bandwidth between various protocols. To support ETS, the switch accepts the ETS traffic class group and bandwidth information Application Priority TLV from auto-upstream devices and propagates it to auto-downstream devices. The switch supports the reception and propagation of ETS information in the automatic configuration port roles. As part of hierarchical scheduling, bandwidth allocation and traffic class groups can be configured by ETS TLVs.

9.5.1. ETS Operation and Dependencies

Using priority-based processing and bandwidth allocations, different Traffic Class Groups (TCGs) within different types of traffic such as LAN, SAN and Management can be configured to provide bandwidth allocation or best effort transmit characteristics.

For ETS to be operational, the following dependency the following three configuration steps need to occur:

1. Configure COS queues to Traffic Class Group mapping for the egress ports.

2. Configure weight percentage (bandwidth allocation) for each TCG.

3. Enable appropriate scheduling algorithm for each TCG.

CoS information is exchanged with peer DCBX devices using ETS TLVs. As part of the ETS TLV, by default, DCBX advertises the following parameters, which are populated on per port basis.

- Mapping between ingress ports 802.1p priority to Traffic Class Group (TCG).
- Bandwidth percentage (weight percentage) of each Traffic Class Group.
- Scheduling algorithm for each Traffic Class Group.

The mapping between the ingress ports 802.1p priority and TCG is not direct. The mapping depends upon:

- The COS map defining the COS queue that a packet is egress forwarded for the ingress 802.1p priority.
- Traffic Class Group map defining the COS queue to TCG mapping.

The indirect mapping between the 802.1p priorities and the associated TCG mapping is advertised by DCBX as part of the ETS TLVs. For this indirect mapping to be valid, the following two parameters must be configured (in addition to the configuration of the TCGs):

1. Configure 8021.p priority to COS mapping for the ingress ports.

2. Enable Trust mode on the ingress ports to trust the 802.1p priority present in the frames.

9.6. VXLAN Gateway Operation and Configuration

9.6.1. Overview

Logically segregated virtual networks in a data center are sometimes referred to as data center VPNs. VXLAN is one of VPNs. Others include E-VPNs, IP VPNs, TRILL, and VPLS.

The encapsulation and decapsulation required by VXLAN is done by devices called Virtual Tunnel Endpoints (VTEPs) or NVEs. VTEPs/NVEs are most commonly implemented within a virtualized server. However, there are cases where it is necessary to implement the VTEP/NVE in a stand-alone networking device. This section describes the functional behavior of a hardware-based VXLAN gateway service and provides configuration scenarios.

9.6.1.1. VXLAN

VXLAN is one method of creating tenant networks on a common network infrastructure. VXLAN encapsulates Ethernet frames in IP packets, thus enabling the network to provide the illusion that hosts connected to

arbitrary access routers are attached to a common layer-2 networks. The VXLAN encapsulation includes a 24-bit virtual network ID (VNID). Hosts can be associated to a VNID and restricted to communicate only with hosts associated to the same VNID. This association segregates communities of interest, or tenants, into different virtual networks. VXLAN allows a public or private data center operator to use a common network infrastructure to provide virtual private network service to multiple tenants while distributing any given tenant's compute and storage resources anywhere in the network infrastructure.

In a data center, VXLAN encapsulation and decapsulation of tenant packets is normally done by a virtual switch within a virtualized server; however, not all tenant systems are virtualized. Non-virtualized tenant systems can participate in a VXLAN by using a VXLAN gateway. A VXLAN gateway is a networking device that does VXLAN encapsulation and decapsulation. A server's first-hop router, often referred to as a top-of-rack (ToR) device, can be a VXLAN gateway.

With VXLAN, the inner Ethernet header can optionally include an incoming VLAN tag. The VXLAN application always strips the inner VLAN information from the incoming Ethernet packet during encapsulation. The inner payload in the VXLAN encapsulated packet does not contain the incoming VLAN tag information in it, which enables flexibility in mapping available VLANs to VNIDs.

The allowed range of VNID values is 1–16777214. VNID 16777215 is reserved for internal purposes.

9.6.2. Functional Description

9.6.2.1. VTEP to VN Association

The operator must configure switches that are to serve as VXLAN gateways. A gateway may serve one or more VPNs. For VXLAN, the operator specifies the virtual network ID (VNID), the type of network (VXLAN), and a method for identifying which incoming native packets belong to the VPN. The ingress VLAN ID can be used as this classifier. Only one VLAN ID can be associated with a specific VNID on a given router. However, the VLAN ID used has no significance beyond that router, and so the same ID can be used on other routers. In this case the number of tenant networks is not limited to VLAN ID space (i.e., 4096). All ingress ports that are members of specified VLAN ID are treated as access ports for the VPN identified by VNID. This defines the access port set for the specified VPN. The access port set for the VXLAN can be altered by updating the VLAN membership configuration. All incoming VLAN traffic is translated to virtual network traffic identified by VNID. A VLAN ID that is already used or configured for routing is not allowed to be configured as an access VLAN for VXLAN.

A source IP address (local VTEP) must be specified for configured VXLAN. The valid source IP interface is either a loopback interface or a routing interface (port-based or VLAN-based) on the router. It is recommended that a loopback interface be dedicated for VXLAN gateway purposes and configured with the intended source IP configuration before associating it with VXLAN. If the configured source IP interface is down or has no IP address, all remote VTEPs in the VPN are considered unreachable. No traffic flows to the remote VTEPs.

Note that the configured source IP address must correspond to an IP address configured on each remote VTEP. Otherwise, the remote VTEPs will discard the gateway's packets.

9.6.2.2. Configuration of Remote VTEPs

Each gateway VTEP must know the set of VTEPs other than itself in VXLAN. This knowledge is necessary because tenant systems can send broadcast and multicast Ethernet frames. For example, ARP requests are generally broadcast. Also, a VTEP may receive a packet for a destination MAC address it has not learned yet. Such a packet is called an unknown frame. The VTEP must send the packet to all other remote VTEPs configured in VXLAN, since the destination may be accessed through any one of them.

VXLAN handle broadcast, multicast, and unknown frames by encapsulating the packet in an IP packet whose destination IP address is an IP multicast group configured for the VN. Each VTEP sends Join messages to join the VN's multicast group. There can be difficulties in using IP multicast to deliver broadcast and unknown frames, the main difficulty being that the data center networks that would be used as underlays often do not enable IP multicast because it does not scale to the size of large public cloud networks. Because of this limitation, VXLAN implementation requires user configuration of the remote VTEPs associated with a particular VPN.

Dynamic VTEP learning through IP multicast is not currently supported.

When a gateway receives a broadcast, multicast, or unknown packet on an access port, it makes a copy of each packet for each of the other VTEP's in the VN, setting the outer IP address to the unicast IP address of the remote VTEP, and setting the outer MAC address to the unicast MAC address of the next hop to the VTEP. The hardware does this packet replication. In this mode, the gateway can still learn L2 entries from packets it decapsulates and, thus, is able to unicast to a single VTEP most of the time.

For each remote VTEP, the operator must specify the following parameters:

- The associated virtual network (specified by VNID).
- The VTEP's IP address. This address is an IP address in the underlay.

The source IP address is inherited from the VXLAN configuration. The system creates overlay tunnels to all configured remote VTEPs in hardware as they become reachable. The system removes the tunnel configuration from hardware when the VTEPs are not reachable.

VXLAN with matching tunnel configuration (i.e., a pair of VTEPs {source or gateway IP address, remote VTEP IP address}) share the same hardware tunnel. Each hardware tunnel has unicast packet and unicast byte counters in either direction (Tx/Rx). When the tunnel is removed from hardware, counters are reset to 0.

If the gateway receives a packet for an unknown VNID or for a known VNID from a VTEP IP address that has not been configured, the gateway drops the packet.

9.6.2.3. VTEP Next-hop Resolution

A remote VTEP is considered reachable if the gateway has a non-default route to the VTEP's IP address. The VXLAN application determines the reachability of the VTEP's address and registers with the routing table manager for changes in the route to that IP address. When there is a route to the VTEP, the VXLAN application copies the next hops of the best route and uses them as the next hop for the packets forwarded to that VTEP. The VXLAN application creates a tunnel in the hardware for each reachable VTEP. The gateway may use multiple next hops to a VTEP, hashing a given flow to an individual next hop as is done in layer-3 routing. The

number of next hops to a VTEP and, thus, the number of next hops for a tunnel, is limited only by the ECMP limit of the switch (or the active SDM template).

The VXLAN application registers with the routing table manager for next-hop resolution changes for each VTEP's remote IP address. When VXLAN receives a next-hop resolution change event, it queries the routing table manager for the new best route and updates the set of next hops to the VTEP. If the VTEP is unreachable, VXLAN deletes the corresponding tunnel in the hardware.

A VTEP cannot be resolved by a default route. The presence of a default route does not provide any confidence that the VTEP is actually reachable.

9.6.2.4. VXLAN UDP Destination Port

The VXLAN standard defines 4789 as the standard UDP destination port to be used for encapsulation and termination. Switches that supported earlier draft versions used custom defined UDP port numbers. To be compatible with those switches, VXLAN supports switch-level VXLAN UDP destination port configuration. By default, the VXLAN UDP destination port is set to 4789 on the switch. The switch terminates incoming VXLAN traffic when the UDP destination port in the VXLAN header matches 4789 and encapsulates VXLAN tenant traffic by putting 4789 in the UDP destination port field in the VXLAN frame.

Users can modify how VXLANs are terminated or encapsulated by changing the default VXLAN UDP destination port configuration on the switch. When the VXLAN UDP destination port is modified, all existing tunnels are modified in the hardware to encapsulate using new VXLAN UDP destination port information. The switch is also configured to terminate VXLAN traffic using the new configuration. There is no or very minimal traffic disruption during this operation.

Note: By default, the switch is configured to generate a source port (in the outer UDP header of the VXLAN frame) that is a hash of the inner Ethernet frame's headers. This is to enable a level of entropy for ECMP/load balancing of the VM to VM traffic across the VXLAN overlay.

9.6.2.5. Tunnels

The VXLAN application creates a tunnel in hardware for each configured and reachable remote VTEP. To create a tunnel in hardware, the application must provide the following tunnel parameters:

- A local IP address. This is the source IP address configured for the VXLAN. The hardware sets the source IP address of the outer IPv4 header to this value.
- The remote IP address. This is the IP address of the VTEP. The hardware sets the destination IP address of the outer IPv4 header to this value.
- A local MAC address, which the hardware uses as the outer source MAC address when encapsulating and sending packets on the tunnel. This MAC address is the MAC address of the originating local routing interface MAC address.
- For VXLAN tunnel, UDP destination port to use in VXLAN header while encapsulation.
- The tunnel VLAN ID. This is the VLAN associated with the outgoing interface in the underlay. If the outgoing interface is a port-based routing interface, this is the VLAN ID assigned internally to the port-

based routing interface. If the outgoing interface is a VLAN routing interface, the tunnel VLAN ID is set to the VLAN ID of this routing interface.

- The next hops in the underlay network. Each next hop is specified as the combination of the following parameters:

  – The internal interface number of the outgoing routing interface in the underlay network.

  – The MAC addresses corresponding to the next hop IP address. The hardware uses this as the destination MAC address of the outer Ethernet header.

9.6.2.6. MAC Learning and Aging

The hardware does MAC learning for VXLAN. Normal MAC learning associates a MAC address with a VLAN and interface. For VXLAN, the hardware learns MAC entries associated with both access ports and network ports. The forwarding entries are learned in the VPN. The VLAN ID field in the entry is replaced by a VPN field. For network-side entries associated with VTEPs, the interface is the hardware tunnel identifier. The MAC address in network-side entries is the MAC address of a tenant system behind a remote VTEP. For access-side entries, the associated interface is the physical or Port-channel interface who are members of the configured VXLAN VLAN. The MAC address in access-side entries is the MAC address of a tenant system behind the local interface (physical or Port-channel interface).

VXLAN MAC entries are not listed in the **show mac-addr-table** command output. They can be listed using **show vxlan address-table**. Both access and network-side entries are listed in the show command output.

The maximum age of a VXLAN MAC entry is the same as normal L2 entries. The user cannot configure a different maximum age for VXLAN MAC entries than for normal L2 entries.

VXLAN performs aging of learned entries in software when the MLAG feature is present in the build package. VXLAN handles entries those are learned in configured VPNs only. It would not handle MAC entries learned in VLANs or listed in the **show mac-addr-table** command output. For packages without the MLAG component, VXLAN relies on hardware aging for MAC entries learned in configured VPNs.

9.6.2.7. Host Configuration

An operator may wish to statically configure host MAC-to-VTEP mappings. Doing so eliminates the initial flooding of packets on all tunnels when the MAC-to-VTEP mapping is unknown. So for each remote VTEP, an operator can optionally configure the MAC addresses of the tenant systems reachable through the VTEP. The maximum allowed static host MAC-to-VTEP binding (or remote tenant systems MAC entries) per tenant is 600. Once this limit is reached, configuring new MAC-to-VTEP bindings for the tenant results in failure. The system generates a log message that describes the reason for failure.

Overall, the system has a maximum allowed limit of 4096 static host MAC-to-VTEP bindings. At any point in time, the sum of all tenants static host MAC-to-VTEP mappings must be less than or equal to the system limit. Once this limit is reached, configuring new MAC-to-VTEP bindings for any tenant results in failure and a log message is generated.

The operator may optionally configure host MAC-to-access port entries as well. The maximum allowed static host MAC-to-interface bindings (or local tenant system MAC entries) per interface (physical or Port-channel)

is 24. Once this limit is reached, configuring new MAC-to-interface bindings for any tenant results in failure and a log message is generated.

9.6.2.8. ECMP

A tunnel may have multiple next hops when the underlay has multiple next hops to the tunnel's remote endpoint. Many data center designs make heavy use of ECMP. To get good traffic distribution within the underlay, it is important that encapsulated packets hash well.

VXLAN encapsulation includes a UDP header. Switches can include the source and destination UDP port in ECMP hash computations. The hardware offers an option for the source VTEP to set the source UDP port to a variable value (hash based on incoming packet Ethernet header) to ensure good ECMP hashing. VXLAN enables this option in hardware by default.

Note: At VXLAN initiation, payload fields are used for hashing at the egress and also to generate the entropy into the UDP source port which becomes part of VXLAN tunnel information. This UDP source port can be used by transit switches for hashing purposes.

9.6.2.9. MTU

VXLAN encapsulation adds 50 bytes of overhead. This additional overhead can cause an encapsulated packet to exceed the MTU of the outgoing port. The gateway does no IP fragmentation while tunneling a packet and is by default configured to set DF=1 in the outer IPv4 header. If an encapsulated packet exceeds the L2 MTU of the outgoing port, the hardware drops it. To avoid this problem, operators must ensure that the L2 MTU on gateway ports to the underlay and underlay network be configured at least 50 bytes larger (for VXLAN) than the MTU on ports on the access side.

The hardware may also enforce an IP MTU. In most cases, network-side ports will be configured as port-based routing interfaces. The IP MTU of these routing interfaces will automatically be adjusted to match the L2 MTU. Therefore, if you adjust the L2 MTU as described above, the hardware should not drop packets because of an IP MTU limitation. If, however, network-side ports are VLAN routing interfaces, you must also increase the IP MTU on each network-side routing interface.

9.6.2.10. TTL and DSCP/TOS

By default, the switch is configured to behave as follows:

- The TTL in the outer IPv4 header during tunnel encapsulation is set to 255.
- For incoming IPv4 packets, the DSCP/TOS value from the incoming IPv4 header is copied into the outer IPv4 header’s DCSP/TOS field during encapsulation. Otherwise, the DSCP/TOS value is set to 0.

9.6.2.11. Packet Forwarding

The gateway forwards all packets in hardware. There is no software forwarding.

9.6.3. VXLAN Configuration Examples

With virtualization technology emerging such as VMware's vSphere, it provides the ability to generate virtual machines (VMs) in a server. A powerful server is able to provide lots of VMs service. When VMs and server grow, VM mobility domain is limited by IP subnetting becomes an issue in data center. VXLAN is layer 2 tunneling feature overcomes IP subnetting limitations. It enables IT administrator to move VMs to any server in the data center, regardless of the data center's subnetting scheme. VXLAN provides the ability as Layer 2 VLAN feature but also be able to provide connectivity extension across the Layer 3 boundary. It is suitable for the large-scale virtualized and multitenant data center designs over a shared common physical infrastructure.

For crossing the Layer 3 network, VXLAN uses MAC-in-UDP encapsulation scheme. The original Layer 2 frame are added a VXLAN header and is encapsulated in an IP UDP packet. With this MAC-in-UDP encapsulation, VXLAN tunnels Layer 2 network over Layer 3 network.

The difference between VLAN and VXLAN:

- VLANs uses a 12-bit VLAN ID to address Layer2 segments, only 4096 VLANs.
- VXLAN uses a 24-bit segment ID known as the VXLAN network identifier (VNI), totally providing 16 million VXLAN segments to coexist in the same administrative domain.
- When VLAN cooperate with STP for avoiding loop issue, it causes half of network links in a network are blocked.
- VXLAN packet are transferred through L3 network based on MAC-in -UDP encapsulation scheme, it has better utilization of available network paths. It also can take the advantage of ECMP (equal-cost multipath) routing and LACP (Link aggregation) to best the network design.

Feature	802.1Q VLAN	VXLAN
Number of VLAN support	4K VLAN	16+ million VLAN
Network packet size	1522 or jumbo frame	Add 50 bytes for VXLAN header
Multicast requirements	N/A	Need Multicast Support PIM-SM, PIM-DM, Bi-direction Note: The switch supports PIM-SM only.
Routing support	Any 802.1Q capable router or switch	Any VTEP capable router
Crossing Layer 3 boundary	No	Yes
ECMP support	No	Yes

Table 9-4: VLAN and VXLAN Comparison

9.6.3.1. Unicast VXLAN Configuration

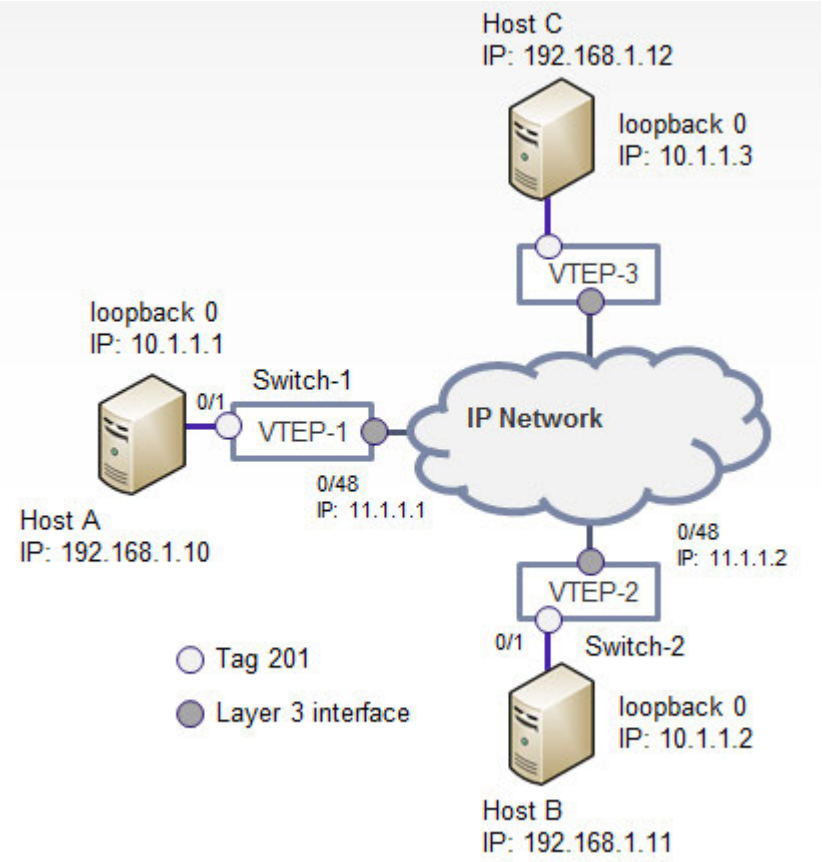


Figure 9-2: Unicast VXLAN Topology

Switch-1 Configuration

Step 1. Create VLAN 201

```
(Switch-1) #configure
(Switch-1) (Config)#vlan database
(Switch-1) (Vlan)#vlan 201
```

Step 2. Enable ip routing

```
(Switch-1) (Config)#ip routing
```

Step 3. Interface Configuration

```
(Switch-1) (Config)#interface loopback 0
(Switch-1) (Interface loopback 0)#ip address 10.1.1.1 255.255.255.255
(Switch-1) (Interface loopback 0)#exit
```

```
(Switch-1) (Config)#interface 0/1
(Switch-1) (Interface 0/1)#switchport allowed vlan add 201
(Switch-1) (Interface 0/1)#switchport tagging 201
(Switch-1) (Interface 0/1)#exit
(Switch-1) (Config)#interface 0/48
(Switch-1) (Interface 0/48)#routing
(Switch-1) (Interface 0/48)#ip address 11.1.1.1 255.255.255.252
(Switch-1) (Interface 0/48)#exit
```

Step 4. Enable OSPF and add network

```
(Switch-1) #configure
(Switch-1) (Config)#router ospf
(Switch-1) (config-router)#router-id 10.1.1.1
(Switch-1) (config-router)#network 10.1.1.1 0.0.0.0 area 0
(Switch-1) (config-router)#network 11.1.1.0 0.0.0.3 area 0
(Switch-1) (config-router)#exit
```

Step 5. Enable VXLAN and configure static VXLAN unicast group

```
(Switch-1) (Config)#interface vxlan 1
(Switch-1) (if-vxlan-1)#vxlan mode unicast
(Switch-1) (if-vxlan-1)#vxlan source-interface loopback 0
(Switch-1) (if-vxlan-1)#vxlan vlan 201 vni 201
(Switch-1) (if-vxlan-1)#vxlan unicast-group 10.1.1.2
(Switch-1) (if-vxlan-1)#vxlan unicast-group 10.1.1.3
(Switch-1) (if-vxlan-1)#exit
```

Switch-2 Configuration

Step 1. Create VLAN 201

```
(Switch-2) #config
(Switch-2) (Config)#vlan database
(Switch-2) (Vlan)#vlan 201
```

Step 2. Enable ip routing

```
(Switch-2) (Config)#ip routing
```

Step 3. Interface Configuration

```
(Switch-2) (Config)#interface loopback 0
(Switch-2) (Interface loopback 0)#ip address 10.1.1.2 255.255.255.255
(Switch-2) (Interface loopback 0)#exit
(Switch-2) (Config)#interface 0/1
```

```
(Switch-2) (Interface 0/1)#switchport allowed vlan add 201
(Switch-2) (Interface 0/1)#switchport tagging 201
(Switch-2) (Interface 0/1)#exit
(Switch-2) (Config)#interface 0/48
(Switch-2) (Interface 0/48)#routing
(Switch-2) (Interface 0/48)#ip address 11.1.1.2 255.255.255.252
(Switch-2) (Interface 0/48)#exit
```

Step 4. Enable OSPF and add network

```
(Switch-2) #configure
(Switch-2) (Config)#router ospf
(Switch-2) (config-router)#router-id 10.1.1.2
(Switch-2) (config-router)#network 10.1.1.2 0.0.0.0 area 0
(Switch-2) (config-router)#network 11.1.1.0 0.0.0.3 area 0
(Switch-2) (config-router)#exit
```

Step 5. Enable VXLAN and configure static VXLAN unicast group

```
(Switch-1) (Config)#interface vxlan 1
(Switch-1) (if-vxlan-1)#vxlan mode unicast
(Switch-1) (if-vxlan-1)#vxlan source-interface loopback 0
(Switch-1) (if-vxlan-1)#vxlan vlan 201 vni 201
(Switch-1) (if-vxlan-1)#vxlan unicast-group 10.1.1.2
(Switch-1) (if-vxlan-1)#vxlan unicast-group 10.1.1.3
(Switch-1) (if-vxlan-1)#exit
```

VXLAN Configuration Verification

Verify VXLAN configuration

```
(Switch-1) #show vxlan


VXLAN Configuration
Interface..................................... Vxlan-1
Mode.......................................... Unicast
UDP Destination Port.......................... 4789
Source Interface.............................. lb0
VXLAN and VLAN Mapping........................ VXLAN ID:201      VLAN ID:201
Unicast Group Address......................... 10.1.1.3
                                   10.1.1.2
```

show remote VTEP learning status

```
(Switch-1) #show vxlan vtep
Remote VTEPs for Vxlan:
10.1.1.2
```

Check the VXLAN address table

```
(Switch-1) #show vxlan address-table
Tenant ID Tenant MAC        VTEP             Interface  AppIfIndex  Entry Type
--------- ----------------- ---------------- ---------- ----------  ----------
201       00:00:00:00:00:0A                   0/1        8529        Learned
201       00:00:00:00:00:0B  10.1.1.2                    338         Learned
```

Appendix A: Term and Acronyms

Table 9-5: Terms and Acronyms

Term	Definition
Access port	A port where native (i.e. unencapsulated) packets are associated with a DCVPN. May be a physical port or a Port-channel.
ACL	Access Control List
Adj-RIB-In	The collection of routing information received from peers
AS	Autonomous System
BFD	Bidirectional Forwarding Detection
BGP	Border Gateway Protocol
BPDU	Bridge Protocol Data Unit
CBS	Committed Burst Size
CIR	Committed Information Rate
CLI	Command Line Interface
CN	Congestion Notification, IEEE 802.1Qau
CoA	Change of Authorization
CoS	Class of Service
CS	Class Selector (as in PHB)
DAC	Dynamic Authorization Client
DAS	Dynamic Authorization Server
DCB	Data Center Bridging
DCPDP	Dual Control Plane Detection Protocol
Default Router	The legacy router. When the Virtual Routing feature is disabled only the Default Router is operational. When the Virtual Routing feature is enabled the Default Router supports all routing protocols and features, while the Virtual Routers support only a subset of features. Also the default router is configured via CLI without specifying the "vrf" keyword.
802.3ad	IEEE Std for Link Aggregation
DSCP	Differentiated Services Code Point
eBGP	Exterior Border Gateway Protocol
ECMP	Equal-Cost Multipath
ECN	Explicit Congestion Notification
ENode	FCoE End Node
ETS	Enhanced Transmission Selection, IEEE 802.1Qaz
FC	Fibre Channel
FCF	FCoE Forwarder
FCoE	Fibre Channel Over Ethernet
FDB	Forwarding Database
FIP	Fibre Channel Initialization Protocol
iBGP	Interior Border Gateway Protocol
IETF	Internet Engineering Task Force
IGMP	Internet Group Management Protocol
IP	Internet Protocol
IP Interface	An interface configured as an IP interface rather than a layer 2 switching interface. An IP interface must be assigned one more IP addresses.
LACP	Link Aggregation Control Protocol
LAG	Link aggregation
LFDB	Label Forwarding Database
LSP	Label Switched Path

Term	Definition
MAC	Media Access Control
MFDB	Multicast Forwarding Database
MIB	Management Information Base
MLAG	Multi Chassis Link Aggregation Group
MLAG partner switch	DUT that is MLAG unaware and forms one end of the Port-channel (with MLAG aware switches on the other end)
MLAG peer switches	DUTs that are MLAG aware and pair to form one end of the Port-channel
MLAG peer-link	Peer-Link between two MLAG peer switches
MAB	MAC Authentication Bypass. This feature provides 802.1x-unaware clients (such as printers and fax machines) controlled access to the network using the devices' MAC address as an identifier.
NAS	Network Access Server
Network port (in VXLAN)	A port where VXLAN tunnels originate or terminate.
Non-redundant ports	Ports on the MLAG aware switch that do not participate in MLAG.
NSF	Non-stop forwarding
NVE	Network Virtualization Edge. NVGRE term for a device or software module that bridges between the overlay and underlay networks. Synonym for VTEP.
PBS	Peak Burst Size
PDU	Protocol data unit
PFC	Priority-based Flow Control,
PIR	Peak Information Rate
QoS	Quality of Service
RADIUS	Remote Authentication Dial In User Services
RED	Random Early Discard
RFC	Request For Comments
Route Leaking	The ability to inject routes belonging to one VR instance into another.
RTO	Routing Table Object. The common routing table, or "RIB", which collects routes from all sources (local, static, dynamic) and determines the most preferred route to each destination.
SDM	Switch Database Management
SNMP	Simple Network Management Protocol
STP	Spanning Tree Protocol
TCP	Transmission Control Protocol
Tenant	An organization for which one or more virtual networks has been provisioned.
Tenant System	A physical or virtual resource, such as a compute or storage device, that is assigned to a specific tenant.
TRILL	Transparent Interconnect of Lots of Links
UDP	User Datagram Protocol
UI	User Interface
Underlay network	IP network that carries tunnel encapsulated traffic from one VTEP/NVE to another.
VLAN	Virtual Local Area Network
VM	Virtual Machine. A virtualized end host.
VN	Virtual Network. The set of tunnels, VTEPs, and tenant systems forming a closed user group. For VXLAN, all traffic in a VN carries the same VNID. This document uses VN interchangeably with DCVPN.
VNID	Virtual network identifier. A 24-bit value that uniquely identifies a VXLAN segment.

Term	Definition
VoIP	Voice over Internet Protocol
VR	Virtual Router
VR-aware	Whether the feature is aware of and works independently in each Virtual Router
VR instance	An instance of the virtual router
VRID	Virtual Router Identifier
VRRP	Virtual Router Redundancy Protocol
VSID	Virtual Segment Subnet IdentifierD ID. A 24-bit value used as a Virtual network identifier in NVGRE.
VTEP	Virtual Tunnel End Point. A device or module that does VXLAN tunnel initiation and termination. Synonym for NVE.
VXLAN	Virtual Extensible Local Area Network
WRED	Weighted Random Early Discard