Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch

Application Guide

For Lenovo Enterprise Network Operating System 8.4

Lenovo™

Note: Before using this information and the product it supports, read the general information in the *Safety information and Environmental Notices and User Guide* documents on the Lenovo *Documentation* CD and the *Warranty Information* document that comes with the product.

Fourth Edition (October 2017)

Contents

Preface

The *Lenovo Flex System Fabric CN4093 10Gb Converged Scalable Switch Application Guide* describes how to configure and use the Enterprise NOS 8.4 software on the Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch (referred to as CN4093 throughout this document).

For documentation about installing the switch physically, see the *Installation Guide* for your CN4093.

Who Should Use This Guide

This guide is intended for network installers and system administrators engaged in configuring and maintaining a network. The administrator should be familiar with Ethernet concepts, IP addressing, Spanning Tree Protocol, and SNMP configuration parameters.

What You'll Find in This Guide

This guide will help you plan, implement, and administer Enterprise NOS software. Where possible, each section provides feature overviews, usage examples, and configuration instructions. The following material is included:

Part 1: Getting Started

This material is intended to help those new to Enterprise NOS products with the basics of switch management. This part includes the following chapters:

- Chapter 1, "Switch Administration," describes how to access the CN4093 in order to configure the switch and view switch information and statistics. This chapter discusses a variety of manual administration interfaces, including local management via the switch console, and remote administration via Telnet, a web browser or via SNMP.
- Chapter 2, "Initial Setup," describes how to use the built-in Setup utility to perform first-time configuration of the switch.
- Chapter 3, "Switch Software Management," describes how to update the N/OS software operating on the switch.

Part 2: Securing the Switch

- Chapter 4, "Securing Administration," describes methods for changing the default switch passwords, using Secure Shell and Secure Copy for administration connections, configuring end-user access control, and placing the switch in protected mode.
- Chapter 5, "Authentication & Authorization Protocols," describes different secure administration for remote administrators. This includes using Remote Authentication Dial-in User Service (RADIUS), as well as TACACS+ and LDAP.

- Chapter 6, "802.1X Port-Based Network Access Control," describes how to authenticate devices attached to a LAN port that has point-to-point connection characteristics. This feature prevents access to ports that fail authentication and authorization and provides security to ports of the CN4093 that connect to blade servers.
- Chapter 7, "Access Control Lists," describes how to use filters to permit or deny specific types of traffic, based on a variety of source, destination, and packet attributes.

Part 3: Switch Basics

- Chapter 8, "VLANs," describes how to configure Virtual Local Area Networks (VLANs) for creating separate network segments, including how to use VLAN tagging for devices that use multiple VLANs. This chapter also describes Protocol-based VLANs, and Private VLANs.
- Chapter 9, "Ports and Link Aggregation (LAG)," describes how to group multiple physical ports together to aggregate the bandwidth between large-scale network devices.
- Chapter 10, "Spanning Tree Protocols," discusses how Spanning Tree Protocol (STP) configures the network so that the switch selects the most efficient path when multiple paths exist. Also includes the Rapid Spanning Tree Protocol (RSTP), Per-VLAN Rapid Spanning Tree Plus (PVRST+), and Multiple Spanning Tree Protocol (MSTP) extensions to STP.
- Chapter 11, "Virtual Link Aggregation Groups," describes using Virtual Link Aggregation Groups (vLAG) to form LAGs spanning multiple vLAG-capable aggregator switches.
- Chapter 12, "Quality of Service," discusses Quality of Service (QoS) features, including IP filtering using Access Control Lists (ACLs), Differentiated Services, and IEEE 802.1p priority values.

Part 4: Advanced Switching Features

- Chapter 13, "Stacking," describes how to implement the stacking feature in the Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch.
- Chapter 14, "Virtualization," provides an overview of allocating resources based on the logical needs of the data center, rather than on the strict, physical nature of components.
- Chapter 15, "Virtual NICs," discusses using virtual NIC (vNIC) technology to divide NICs into multiple logical, independent instances.
- Chapter 16, "VMready," discusses virtual machine (VM) support on the CN4093.
- Chapter 17, "FCoE and CEE," discusses using various Converged Enhanced Ethernet (CEE) features such as Priority-based Flow Control (PFC), Enhanced Transmission Selection (ETS) and FIP Snooping for solutions such as Fibre Channel over Ethernet (FCoE).
- Chapter 18, "Fibre Channel," describes how to configure the CN4093 for use with Fibre Channel networks.

- Chapter 19, "Edge Virtual Bridging (EVB)," discusses the IEEE 802.1Qbg—a standards-based protocol that defines how virtual Ethernet bridges exchange configuration information. EVB bridges the gap between physical and virtual network resources, thus simplifying network management.
- Chapter 20, "Static Multicast ARP," discusses the configuration of a static ARP entry with multicast MAC address for Microsoft's Network Load Balancing (NLB) feature to function efficiently.
- Chapter 21, "Unified Fabric Port," describes how UFP logically subdivides a high-speed physical link connecting to a server NIC or to a Converged Network Adapter (CNA). UFP provides a switch fabric component to control the NIC.
- Chapter 22, "Switch Partition," describes the creation of multiple partitions within a switch to form a virtual switching context with respect to data plane partition of a switch.

Part 5: IP Routing

- Chapter 23, "Basic IP Routing," describes how to configure the CN4093 for IP routing using IP subnets, BOOTP, and DHCP Relay.
- Chapter 24, "Internet Protocol Version 6," describes how to configure the CN4093 for IPv6 host management.
- Chapter 25, "Using IPsec with IPv6," describes how to configure Internet Protocol Security (IPsec) for securing IP communications by authenticating and encrypting IP packets, with emphasis on Internet Key Exchange version 2, and authentication/confidentiality for OSPFv3.
- Chapter 26, "Routing Information Protocol," describes how the Enterprise NOS software implements standard Routing Information Protocol (RIP) for exchanging TCP/IP route information with other routers.
- Chapter 27, "Internet Group Management Protocol," describes how the Enterprise NOS software implements IGMP Snooping or IGMP Relay to conserve bandwidth in a multicast-switching environment.
- Chapter 28, "Multicast Listener Discovery," describes how Multicast Listener Discovery (MLD) is used with IPv6 to support host users requests for multicast data for a multicast group.
- Chapter 29, "Border Gateway Protocol," describes Border Gateway Protocol (BGP) concepts and features supported in Enterprise NOS.
- Chapter 30, "OSPF," describes key Open Shortest Path First (OSPF) concepts and their implemented in Enterprise NOS, and provides examples of how to configure your switch for OSPF support.
- Chapter 31, "Protocol Independent Multicast," describes how multicast routing can be efficiently accomplished using the Protocol Independent Multicast (PIM) feature.

Part 6: High Availability Fundamentals

- Chapter 32, "Basic Redundancy," describes how the CN4093 supports redundancy through aggregation and Hotlinks.
- Chapter 33, "Layer 2 Failover," describes how the CN4093 supports high-availability network topologies using Layer 2 Failover.

- Chapter 34, "Virtual Router Redundancy Protocol," describes how the CN4093 supports high-availability network topologies using Virtual Router Redundancy Protocol (VRRP).

Part 7: Network Management

- Chapter 35, "Link Layer Discovery Protocol," describes how Link Layer Discovery Protocol helps neighboring network devices learn about each others' ports and capabilities.
- Chapter 36, "Simple Network Management Protocol," describes how to configure the switch for management through an SNMP client.
- Chapter 37, "Service Location Protocol," describes the Service Location Protocol (SLP) that allows the switch to provide dynamic directory services.
- Chapter 38, "System License Keys," describes how to manage Features on Demand (FoD) licenses and how to allocate bandwidth between physical ports within the installed licenses' limitations.

Part 8: Monitoring

- Chapter 40, "Remote Monitoring," describes how to configure the RMON agent on the switch, so that the switch can exchange network monitoring data.
- Chapter 41, "sFLOW, described how to use the embedded sFlow agent for sampling network traffic and providing continuous monitoring information to a central sFlow analyzer.
- Chapter 42, "Port Mirroring," discusses tools how copy selected port traffic to a monitor port for network analysis.

Part 9: Appendices

- Appendix A, "Glossary," describes common terms and concepts used throughout this guide.
- Appendix B, "Getting help and technical assistance," describes how to get help.
- Appendix C, "Notices," provides trademark and other compliance information.

Additional References

Additional information about installing and configuring the CN4093 is available in the following guides:

- *Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch Installation Guide*
- *Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch Command Reference for Lenovo Network Operating System 8.4*
- *Lenovo Network Browser-Based Interface Quick Guide*

Typographic Conventions

The following table describes the typographic styles used in this book.

Table 1. *Typographic Conventions*

Typeface or Symbol	Meaning	Example
`ABC123`	This type is used for names of commands, files, and directories used within the text. It also depicts on-screen computer output and prompts.	View the `readme.txt` file. `Main#`
`ABC123`	This bold type appears in command examples. It shows text that must be typed in exactly as shown.	`Main#` **`sys`**
<ABC123>	This italicized type appears in command examples as a parameter placeholder. Replace the indicated text with the appropriate real name or value when using the command. Do not type the brackets. This also shows book titles, special terms, or words to be emphasized.	To establish a Telnet session, enter: `host#` **`telnet`** *<IP address>* Read your *User's Guide* thoroughly.
`[ ]`	Command items shown inside brackets are optional and can be used or excluded as the situation demands. Do not type the brackets.	`host#` **`ls`** `[`**`-a`**`]`
`|`	The vertical bar (`|`) is used in command examples to separate choices where multiple options exist. Select only one of the listed options. Do not type the vertical bar.	`host#` **`set left|right`**
AaBbCc123	This block type depicts menus, buttons, and other controls that appear in Web browsers and other graphical interfaces.	Click the **Save** button.

Part 1: Getting Started

Chapter 1. Switch Administration

Your CN4093 10 Gb Converged Scalable Switch is ready to perform basic switching functions right out of the box. Some of the more advanced features, however, require some administrative configuration before they can be used effectively.

The extensive Enterprise NOS switching software included in the CN4093 provides a variety of options for accessing the switch to perform configuration, and to view switch information and statistics.

This chapter discusses the various methods that can be used to administer the switch.

Administration Interfaces

The switch software provides a variety of user-interfaces for administration. These interfaces vary in character and in the methods used to access them: some are text-based, and some are graphical; some are available by default, and some require configuration; some can be accessed by local connection to the switch, and others are accessed remotely using various client applications. For example, administration can be performed using any of the following:

- The Flex System chassis management module tools for general chassis management
- A built-in, text-based command-line interface and menu system for access via serial-port connection or an optional Telnet or SSH session
- The built-in Browser-Based Interface (BBI) available using a standard web-browser
- SNMP support for access through network management software such as IBM Director.

The specific interface chosen for an administrative session depends on user preferences, as well as the switch configuration and the available client tools.

In all cases, administration requires that the switch hardware is properly installed and turned on. (see the *Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch Installation Guide*).

Chassis Management Module

The CN4093 10 Gb Converged Scalable Switch is an integral subsystem within the overall Lenovo Flex System. The Flex System chassis also includes a chassis management module (CMM) as the central element for overall chassis management and control. Using the tools available through the CMM, the administrator can configure many of the CN4093 features and can also access other CN4093 administration interfaces.

For more information, see "Using the Chassis Management Module" on page 32.

Industry Standard Command Line Interface

The Industry Standard Command Line Interface (ISCLI) provides a simple, direct method for switch administration. Using a basic terminal, you can issue commands that allow you to view detailed information and statistics about the switch, and to perform any necessary configuration and switch software maintenance.

You can establish a connection to the CLI in any of the following ways:

- Serial connection via the serial port on the CN4093 (this option is always available)
- Telnet connection over the network
- SSH connection over the network

When you first access the switch, you must enter the default username and password: USERID; PASSW0RD (with a zero). You are required to change the password after first login.

Browser-Based Interface

The Browser-based Interface (BBI) provides access to the common configuration, management and operation features of the CN4093 through your Web browser.

For more information, refer to the *Enterprise NOS BBI Quick Guide*.

Establishing a Connection

The factory default settings permit initial switch administration through the built-in serial port, as well as default IP addresses on VLAN 1 and the out-of-band management port.

To facilitate switch access, the in-band and out-of-band management interfaces are configured with factory default IP addresses, as follows:

- VLAN 1/Interface 1: `192.168.49.50/24`
- Out-of-band Management Port 1: `192.168.50.50/24`

Remote access using the network requires the accessing terminal to have a valid, routable connection to the switch interface. The client IP address may be configured manually, or an IPv4 address can be provided automatically through the switch using a service such as DHCP or BOOTP relay (see "BOOTP/DHCP Client IP Address Services" on page 40), or an IPv6 address can be obtained using IPv6 stateless address configuration.

Note: Throughout this manual, *IP address* is used in places where either an IPv4 or IPv6 address is allowed. IPv4 addresses are entered in dotted-decimal notation (for example, 10.10.10.1), while IPv6 addresses are entered in hexadecimal notation (for example, 2001:db8:85a3::8a2e:370:7334). In places where only one type of address is allowed, *IPv4 address* or *IPv6 address* is specified.

Using the Chassis Management Module

The CN4093 is an integral subsystem within the overall Lenovo Flex System. The Flex System chassis includes a chassis management module (CMM) as the central element for overall chassis management and control.

The CN4093 uses port 43 (MGT1) to communicate with the chassis management module(s). Even when the CN4093 is in a factory default configuration, you can use the 1Gb Ethernet port on each CMM to configure and manage the CN4093.

For more information about using the chassis management module, see the *Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch Installation Guide*.

Factory-Default vs. CMM-Assigned IP Addresses

Each CN4093 must be assigned its own Internet Protocol version 4 (IPv4) address, which is used for communication with an SNMP network manager or other transmission control protocol/Internet Protocol (TCP/IP) applications (for example, BOOTP or TFTP). The factory-default IPv4 address is 10.90.90.*x*, where *x* is based on the number of the bay into which the CN4093 is installed. For additional information, see the *Installation Guide*. The chassis management module assigns an IPv4 address of 192.168.70.1*xx*, where *xx* is also based on the number of the bay into which each CN4093 is installed, as shown in the following table:

Table 2. *CN4093 IPv4 addresses, by switch-module bay numbers*

Bay Number	Factory-Default IPv4 Address	IPv4 Address Assigned by CMM
Bay 1	10.90.90.91	192.168.70.120
Bay 2	10.90.90.92	192.168.70.121

Table 2. *CN4093 IPv4 addresses, by switch-module bay numbers*

Bay Number	Factory-Default IPv4 Address	IPv4 Address Assigned by CMM
Bay 3	10.90.90.93	192.168.70.122
Bay 4	10.90.90.94	192.168.70.123

Note: CN4093s installed in Bay 1 and Bay 2 connect to server NICs 1 and 2, respectively.

Using Telnet

A Telnet connection offers the convenience of accessing the switch from a workstation connected to the network. Telnet access provides the same options for user and administrator access as those available through the console port.

By default, Telnet access is disabled. Use the following commands (available on the console only) to enable or disable Telnet access:

```
CN 4093(config)# [no] access telnet enable
```

Once the switch is configured with an IP address and gateway, you can use Telnet to access switch administration from any workstation connected to the management network.

To establish a Telnet connection with the switch, run the Telnet program on your workstation and issue the following Telnet command:

```
telnet <switch IPv4 or IPv6 address>
```

You will then be prompted to enter a password as explained "Switch Login Levels" on page 47.

Two attempts are allowed to log in to the switch. After the second unsuccessful attempt, the Telnet client is disconnected via TCP session closure.

Using Secure Shell

Although a remote network administrator can manage the configuration of a CN4093 via Telnet, this method does not provide a secure connection. The Secure Shell (SSH) protocol enables you to securely log into another device over a network to execute commands remotely. As a secure alternative to using Telnet to manage switch configuration, SSH ensures that all data sent over the network is encrypted and secure.

The switch can do only one session of key/cipher generation at a time. Thus, a SSH/SCP client will not be able to login if the switch is doing key generation at that time. Similarly, the system will fail to do the key generation if a SSH/SCP client is logging in at that time.

The supported SSH encryption and authentication methods are listed below.

- Server Host Authentication: Client RSA-authenticates the switch when starting each connection

- Key Exchange: ecdh-sha2-nistp521, ecdh-sha2-nistp384, ecdh-sha2-nistp256, ecdh-sha2-nistp224, ecdh-sha2-nistp192, rsa2048-sha256, rsa1024-sha1, diffie-hellman-group-exchange-sha256, diffie-hellman-group-exchange-sha1, diffie-hellman-group14-sha1, diffie-hellman-group1-sha1
- Encryption: aes128-ctr, aes128-cbc, rijndael128-cbc, blowfish-cbc,3des-cbc, arcfour256, arcfour128, arcfour
- MAC: hmac-sha1, hmac-sha1-96, hmac-md5, hmac-md5-96
- User Authentication: Local password authentication, RADIUS, TACACS+

The following SSH clients have been tested:

- OpenSSH_5.1p1 Debian-3ubuntu1
- SecureCRT 5.0 (Van Dyke Technologies, Inc.)
- Putty beta 0.60

Note: The Enterprise NOS implementation of SSH supports version 2.0 and supports SSH client version 2.0.

Using SSH with Password Authentication

By default, the SSH feature is enabled. For information about enabling and using SSH for switch access, see "Secure Shell and Secure Copy" on page 91.

Once the IP parameters are configured and the SSH service is enabled, you can access the command line interface using an SSH connection.

To establish an SSH connection with the switch, run the SSH program on your workstation by issuing the SSH command, followed by the switch IPv4 or IPv6 address:

```
ssh <switch IP address>
```

You will then be prompted to enter a password as explained "Switch Login Levels" on page 47.

Using SSH with Public Key Authentication

SSH can also be used for switch authentication based on asymmetric cryptography. Public encryption keys can be uploaded on the switch and used to authenticate incoming login attempts based on the clients' private encryption key pairs. After a predefined number of failed public key login attempts, the switch reverts to password-based authentication.

To set up public key authentication:

1. Enable SSH:

```
CN 4093(config)# ssh enable
```

2. Import the public key file using SFTP or TFTP for the admin user account:

```
CN 4093(config)# copy {sftp|tftp} public-key
Port type ["DATA"/"MGT"]: mgt
Address or name of remote host: 9.43.101.151
Source file name: 11.key
Username of the public key: admin
Confirm download operation (y/n) ? y
```

Note: When prompted to input a username, a valid user account name must be entered. If no username is entered, the key is stored on the switch, and can be assigned to a user account later.

Note: A user account can have up to 100 public keys set up on the switch.

3. Configure a maximum number of 3 failed public key authentication attempts before the system reverts to password-based authentication:

```
CN 4093(config)# ssh maxauthattempts 3
```

Once the public key is configured on the switch, the client can use SSH to login from a system where the private key pair is set up:

```
ssh <switch IP address>
```

Using a Web Browser

The switch provides a Browser-Based Interface (BBI) for accessing the common configuration, management and operation features of the CN4093 through your Web browser.

You can access the BBI directly from an open Web browser window. Enter the URL using the IP address of the switch interface (for example, `http://`*<IPv4 or IPv6 address>*).

When you first access the switch, you must enter the default username and password: USERID; PASSW0RD (with a zero). You are required to change the password after first login.

Configuring HTTP Access to the BBI

By default, BBI access via HTTP is disabled on the switch.

To enable or disable HTTP access to the switch BBI, use the following commands:

```
CN 4093(config)# access http enable          (Enable HTTP access)
  -or-
CN 4093(config)# no access http enable       (Disable HTTP access)
```

The default HTTP web server port to access the BBI is port 80. However, you can change the default Web server port with the following command:

```
CN 4093(config)# access http port <TCP port number>
```

To access the BBI from a workstation, open a Web browser window and type in the URL using the IP address of the switch interface (for example, `http://`*<IPv4 or IPv6 address>*).

Configuring HTTPS Access to the BBI

The BBI can also be accessed via a secure HTTPS connection.

1. Enable HTTPS.

   By default, BBI access via HTTPS is enabled on the switch. To disable or re-enable BBI access via HTTPS, use the following commands:

```
CN 4093(config)# no access https enable        (Disable HTTPS access)
   -or-
CN 4093(config)# access https enable           (Enable HTTPS access)
```

2. Set the HTTPS server port number (optional).

   The default HTTPS web server port to access the BBI is port 443. However, you can change the default Web server port with the following command:

```
CN 4093(config)# access https port <x>
```

3. Generate the HTTPS certificate.

   Accessing the BBI via HTTPS requires that you generate a certificate to be used during the key exchange. A default certificate is created the first time HTTPS is enabled, but you can create a new certificate defining the information you want to be used in the various fields.

```
CN 4093(config)# access https generate-certificate
Country Name (2 letter code) []:  <country code>
State or Province Name (full name) []:  <state>
Locality Name (eg, city) []:  <city>
Organization Name (eg, company) []:  <company>
Organizational Unit Name (eg, section) []:  <org. unit>
Common Name (eg, YOUR name) []:  <name>
Email (eg, email address) []:  <email address>
Confirm generating certificate? [y/n]: y
Generating certificate. Please wait (approx 30 seconds)
restarting SSL agent
```

4. Save the HTTPS certificate.

   The certificate is valid only until the switch is rebooted. To save the certificate so that it is retained beyond reboot or power cycles, use the following command:

```
CN 4093(config)# access https save-certificate
```

   When a client (such as a web browser) connects to the switch, the client is asked to accept the certificate and verify that the fields match what is expected. Once BBI access is granted to the client, the BBI can be used as described in the *Enterprise NOS BBI Quick Guide*.

BBI Summary

The BBI is organized at a high level as follows:

Context buttons—These buttons allow you to select the type of action you wish to perform. The *Configuration* button provides access to the configuration elements for the entire switch. The *Statistics* button provides access to the switch statistics and state information. The *Dashboard* button allows you to display the settings and operating status of a variety of switch features.

Navigation Window—This window provides a menu list of switch features and functions:

- **System**—this folder provides access to the configuration elements for the entire switch.
- **Switch Ports**—Configure each of the physical ports on the switch.
- **Port-Based Port Mirroring**—Configure port mirroring behavior.
- **Layer 2**—Configure Layer 2 features for the switch.
- **RMON Menu**—Configure Remote Monitoring features for the switch.
- **Layer 3**—Configure Layer 3 features for the switch.
- **QoS**—Configure Quality of Service features for the switch.
- **Access Control**—Configure Access Control Lists to filter IP packets.
- **Virtualization –** Configure VMready for virtual machine (VM) support.

For information on using the BBI, refer to the *Enterprise NOS BBI Quick Guide*.

Using Simple Network Management Protocol

Enterprise NOS provides Simple Network Management Protocol (SNMP) version 1, version 2, and version 3 support for access through any network management software, such as IBM Director.

Note: SNMP is disabled by default. However, if community strings are already configured on the switch, any software update will leave SNMP enabled.

To access the SNMP agent on the CN4093, the read and write community strings on the SNMP manager must be configured to match those on the switch.

The read and write community strings on the switch can be changed using the following commands:

```
CN 4093(config)# snmp-server read-community <1-32 characters>
CN 4093(config)# snmp-server write-community <1-32 characters>
```

The SNMP manager can reach any of the IP interfaces on the switch.

For the SNMP manager to receive the SNMPv1 traps sent out by the SNMP agent on the switch, configure the trap host on the switch with the following commands:

```
CN 4093(config)# snmp-server trap-source <trap source IP interface>
CN 4093(config)# snmp-server host <IPv4 address> <trap host community string>
```

For more information on SNMP usage and configuration, see "Simple Network Management Protocol" on page 559.

BOOTP/DHCP Client IP Address Services

For remote switch administration, the client terminal device must have a valid IP address on the same network as a switch interface. The IP address on the client device may be configured manually, or obtained automatically using IPv6 stateless address configuration, or an IPv4 address may obtained automatically via BOOTP or DHCP relay as discussed below.

The CN4093 can function as a relay agent for Bootstrap Protocol (BOOTP) or DHCP. This allows clients to be assigned an IPv4 address for a finite lease period, reassigning freed addresses later to other clients.

Acting as a relay agent, the switch can forward a client's IPv4 address request to up to five BOOTP/DHCP servers. In addition to the five global BOOTP/DHCP servers, up to five domain-specific BOOTP/DHCP servers can be configured for each of up to 10 VLANs.

When a switch receives a BOOTP/DHCP request from a client seeking an IPv4 address, the switch acts as a proxy for the client. The request is forwarded as a UDP Unicast MAC layer message to the BOOTP/DHCP servers configured for the client's VLAN, or to the global BOOTP/DHCP servers if no domain-specific BOOTP/DHCP servers are configured for the client's VLAN. The servers respond to the switch with a Unicast reply that contains the IPv4 default gateway and the IPv4 address for the client. The switch then forwards this reply back to the client.

DHCP is described in RFC 2131, and the DHCP relay agent supported on the CN4093 is described in RFC 1542. DHCP uses UDP as its transport protocol. The client sends messages to the server on port 67 and the server sends messages to the client on port 68.

BOOTP and DHCP relay are collectively configured using the BOOTP commands and menus on the CN4093.

Host Name Configuration

The CN4093 supports DHCP host name configuration as described in RFC 2132, option 12. DHCP host name configuration is enabled by default.

Host name can be manually configured using the following command:

```
CN 4093(config)# hostname <name>
```

If the host name is manually configured, the switch does not replace it with the host name received from the DHCP server.

After the host name is configured on the switch, if DHCP or DHCP host name configuration is disabled, the switch retains the host name.

The switch prompt displays the host name.

Host name configuration can be enabled/disabled using the following command:

```
CN 4093(config)# [no] system dhcp hostname
```

SYSLOG Server

During switch startup, if the switch fails to get the configuration file, a message can be recorded in the SYSLOG server.

The CN4093 supports requesting of a SYSLOG server IP address from the DHCP server as described in RFC 2132, option 7. DHCP SYSLOG server request option is enabled by default.

Manually configured SYSLOG server takes priority over DHCP SYSLOG server.

Up to two SYSLOG server addresses received from the DHCP server can be used. The SYSLOG server can be learnt over a management port or a data port.

Use the **`show logging`** command to view the SYSLOG server address.

DHCP SYSLOG server address option can be enabled/disabled using the following command:

```
CN 4093(config)# [no] system dhcp syslog
```

DHCP Snooping

DHCP snooping provides security by filtering untrusted DHCP packets and by building and maintaining a DHCP snooping binding table. This feature is applicable only to IPv4 and only works in non-stacking mode.

An untrusted interface is a port that is configured to receive packets from outside the network or firewall. A trusted interface receives packets only from within the network. By default, all DHCP ports are untrusted.

The DHCP snooping binding table contains the MAC address, IP address, lease time, binding type, VLAN number, and port number that correspond to the local untrusted interface on the switch; it does not contain information regarding hosts interconnected with a trusted interface.

By default, DHCP snooping is disabled on all VLANs. You can enable DHCP snooping on one or more VLANs. You must enable DHCP snooping globally. To enable this feature, enter the following commands:

```
CN 4093(config)# ip dhcp snooping vlan <vlan number(s)>
CN 4093(config)# ip dhcp snooping
```

Note: When you make a DHCP release from a client, the switch does not forward the Unicast DHCP release packet to the server, the entry is not removed from the DHCP snooping binding table, and the counter for Received Request packets does not increase even though the release packet does arrive at the switch.

If you want the DHCP Renew/Release packet to be forwarded to the server and the corresponding entry removed from the DHCP snooping binding table, configure an interface IP address with the sam subnet in the same VLAN.

Easy Connect Wizard

Lenovo EasyConnect (EZC) is a feature designed to simplify switch configuration. A set of predefined configurations can be applied on the switch via ISCLI. By launching the EZC Wizard, you are prompted for a minimal set of input and the tool automatically customizes the switch software.

The EZC Wizard allows you to choose one of the following configuration modes:

- **Basic System** mode supports settings for hostname, static management port IP, netmask, and gateway.
- **Transparent** mode collects server and uplink port settings. vNIC groups are used to define the loop free domains.

  Note: You can either accept the static defaults or enter a different port list for uplink and/or server ports.
- **Redundant** mode refers to VLAG settings.

The EZC configuration will be applied immediately. Any existing configuration will be deleted, the current active or running configuration will not be merged or appended to the EZC configuration.

For any custom settings that are not included in the predefined configuration sets, the user has to do it manually.

Notes:

- EZC is not available in stacking mode.
- To support scripting, the feature also has a single-line format. For more information, please refer to *Lenovo Networking ISCLI Reference Guide*.

Note: To support scripting, the feature also has a single-line format. For more information, please refer to *Lenovo Networking ISCLI Reference Guide*.

Configuring the Easy Connect Wizard

To launch the EZC Wizard, use the following command:

```
CN 4093# easyconnect
```

The wizard displays the available predefined configuration modes. You are prompted to select one of the following options:

```
CN 4093# easyconnect
Auto configures the switch into a set configuration based on the input
provided.
Current configuration will be overwritten with auto configuration
settings.
The wizard can be canceled anytime by pressing Ctrl+C.
Select which of the following features you want enabled:
#Configure Basic system (yes/no)?
#Configure Transparent mode (yes/no)?
```

Basic System Mode Configuration Example

This example shows the parameters available for configuration in Basic System mode:

```
CN 4093# easyconnect
Configure Basic system (yes/no)? y

Please enter "none" for no hostname.
Enter hostname(Default: None)? host

Please enter "dhcp" for dhcp IP.
Select management IP address (Current: 10.241.13.32)?
Enter management netmask(Current: 255.255.255.128)?
Enter management gateway:(Current: 10.241.13.1)?

Pending switch port configuration:

    Hostname: host
    Management interface:
        IP:      10.241.13.32
        Netmask: 255.255.255.128
        Gateway: 10.241.13.1
Confirm erasing current config to re-configure Easy Connect (yes/no)?
```

```
CN 4093# easyconnect
Configure Basic system (yes/no)? y

Please enter "none" for no hostname.
Enter hostname(Default: None)? Host

Select management port number(Default: 1)?

Please enter "dhcp" for dhcp IP.
Select management IP address (Current: 10.241.13.32)?
Enter management netmask(Current: 255.255.255.128)?
Enter management gateway:(Current: 10.241.13.1)?

Pending switch port configuration:

    Hostname: host
    Management interface:
        Port:    1
        IP:      10.241.13.32
        Netmask: 255.255.255.128
        Gateway: 10.241.13.1
```

Note: You can either accept the default values or enter new parameters.

Transparent Mode Configuration Example

This example shows the parameters available for configuration in Transparent mode:

```
CN 4093# # easyconnect
Configure Transparent mode (yes/no)? y
Select Uplink Ports (Static Defaults:  17-24)?
The following Uplink ports will be enabled:
        Uplink ports(1G/10G):  17-24
Select Server Ports (Static Defaults:  25-64)?
The following Server ports will be enabled:
        Server ports(1G/10G):  25-64
Pending switch configuration:

    Uplink Ports:    17-24
    Server Ports:    25-64
    Disabled Ports:  1,5,9,13
Confirm erasing current config to re-configure Easy Connect (yes/no)?
```

Notes:

- If your selection for a port group contains ports of different mode or speed, the selection is not valid and you are guided to either select other ports or change the speed of the ports.
- If your selection for an uplink port group contains ports of different mode or speed, the selection is not valid and you are guided to select other ports. Server ports can have ports of different mode or speed selected at the same time.
- You can either accept the static defaults or enter a different port list for uplink and/or server ports.

Redundant Mode Configuration Example

This example shows the parameters available for configuration in Redundant mode:

```
CN 4093# #easyconnect
Configure Switch Redundant mode (yes/no)? y

Note: It is recommended to select Basic system configuration in order to
set the management IP address used for vLAG health check.

Configure Basic system (yes/no)? y

Configure this switch as vLAG Primary or Secondary Peer
(primary/secondary)? prim

The following ISL ports will be enabled:

Select vLAG TierID (Default: 101)?

Select management IP address (Current: 192.168.49.50)?

Enter management netmask (Current: 255.255.255.0)?

Select Peer IP address for vLAG healthcheck (Default: 1.1.1.2)?
Warning: vLAG healthcheck Peer IP is not reachable.
Do you want to select another Peer IP (yes/no)? y
Select Peer IP address for vLAG healthcheck (Default: 1.1.1.2)?
Warning: vLAG healthcheck Peer IP is not reachable.
Do you want to select another Peer IP (yes/no)? n

The following Uplink ports will be enabled:

The following Downlink ports will be enabled:
```

```
Please enter "none" for no hostname.
Enter hostname(Default: Primary VLAG)?

Please enter "none" for no gateway.
Enter management gateway:(Default: 0.0.0.0)?

Pending switch configuration:

    vLAG switch type:   Primary
    ISL Ports:
    vLAG TierID:        101
    vLAG Peer IP:       1.1.1.2
    Uplink Ports:
    Downlink Ports:
    Disabled Ports:      empty

    Hostname: Primary VLAG
 Management interface:
        IP:      192.168.49.50
        Netmask: 255.255.255.0
        Gateway: 0.0.0.0

Confirm erasing current config to re-configure Easy Connect (yes/no)?
```

Notes:

- If your selection for a port group contains ports of different speed, the selection is not valid, and you are guided to either select other ports or change the speed of the ports.
- All unused port are configured as shut down in the configuration dump.
- You can either accept the static defaults or enter a different port list for ISL, uplink, and/or downlink ports.

Switch Login Levels

To enable better switch management and user accountability, three levels or *classes* of user access have been implemented on the CN4093. Levels of access to CLI, Web management functions, and screens increase as needed to perform various switch management tasks. Conceptually, access classes are defined as follows:

- User interaction with the switch is completely passive—nothing can be changed on the CN4093. Users may display information that has no security or privacy implications, such as switch statistics and current operational state information.
- Operators can only effect temporary changes on the CN4093. These changes will be lost when the switch is rebooted/reset. Operators have access to the switch management features used for daily switch operations. Because any changes an operator makes are undone by a reset of the switch, operators cannot severely impact switch operation.
- Administrators are the only ones that may make permanent changes to the switch configuration—changes that are persistent across a reboot/reset of the switch. Administrators can access switch functions to configure and troubleshoot problems on the CN4093. Because administrators can also make temporary (operator-level) changes as well, they must be aware of the interactions between temporary and permanent changes.

Access to switch functions is controlled through the use of unique user names and passwords. Once you are connected to the switch via console, remote Telnet, or SSH, you are prompted to enter a password. The default user names/password for each access level are listed in the following table.

Note: It is recommended that you change default switch passwords after initial configuration and as regularly as required under your network security policies. For more information, see "Changing the Switch Passwords" on page 90.

Table 3. *User Access Levels - Default Settings*

User Account	Password	Description and Tasks Performed	Status
`user`	`user`	The User has no direct responsibility for switch management. He or she can view all switch status information and statistics, but cannot make any configuration changes to the switch.	Disabled
`oper`	`oper`	The Operator manages all functions of the switch. The Operator can reset ports, except the management ports.	Disabled
`admin`	`admin`	The Administrator has complete access to all menus, information and configuration commands on the CN4093, including the ability to change both the user and administrator passwords.	Enabled

Note: Access to each user level (except admin account) can be disabled by setting the password to an empty value. To disable admin account, use the command: `CN 4093(config)# no access user administrator-enable`. Admin account can be disabled only if there is at least one user account enabled and configured with administrator privilege.

Administrator Password Recovery

Follow these steps to reset the password of the `admin` user to the default value:

Note: Password recovery process involves reloading the switch. Make sure to save any recent switch configuration changes before performing these steps.

1. Connect to the switch using the console port.
2. Reload the switch.
3. When the system displays Memory Test, press **<Shift + B>**. The Boot Management menu displays:

```
**** System Reset from boot iscli ****
Disable the Transceivers ...
Unmount the File System ...
Unmounting filesystem
Wait for umount to finish.Done
Waiting for I2C Transactions to Finish ...


U-Boot 2009.06 (Aug 21 2015 - 12:35:27) MPC83XX

Reset Status:

CPU: e300c4, MPC8378A, Rev: 2.1 at 792 MHz, CSB: 396 MHz
Board: Networking OS RackSwitch G8052
I2C: ready
DRAM: 1 GB

Memory Test ..........
```

4. Select **`C - Change configuration block`** from the Boot menu by entering **`C`**. When prompted for the configuration block, enter **`f`**:

```
Boot Menu Mode

Platform: Rack Switch G8052 (version 0.0.0.1)
FLASH: 256 MB
PCIE0: Link

Boot Management Menu
        I - Change booting image
        C - Change configuration block
        R - Boot in recovery mode (xmodem download of images to recover
switch)
        Q - Reboot
        E - Exit
Please choose your menu option: c

Currently using active configuration block
Enter configuration block: a, b or f (active, backup or factory): f
```

5. Enter **Q** to reboot the switch:

```
Boot Management Menu
        I - Change booting image
        C - Change configuration block
        R - Boot in recovery mode (xmodem download of images to recover
switch)
        Q - Reboot
        E - Exit
Please choose your menu option: q
Resetting the board.
```

6. After the reload is complete, log into the switch by using the default user `admin` with the default password `admin`.

7. Enter configuration mode (config). Copy the active configuration to the running configuration by using the **`copy active-config running-config admin-pw-bypass`** command.

```
CN 4093> ena
Enable privilege granted.
CN 4093# configure terminal
Enter configuration commands, one per line. End with Ctrl/Z.
CN 4093(config)# copy active-config running-config admin-pw-bypass
Loading to current configuration.
```

8. Use the **`show run`** command to confirm the configuration is recovered. Set the new admin and enable passwords. Save the running configuration to startup configuration.

```
CN 4093(config)# password
Changing admin password; validation required:
Enter current local admin password:
Enter new admin password (max 64 characters):
Re-enter new admin password:
New admin password accepted.

Password changed and applied, but not saved.
Notifying administrator to save changes.

CN 4093(config)# enable password ?
  WORD The UNENCRYPTED (cleartext) 'enable' password
CN 4093(config)# enable password admin1
CN 4093(config)# copy running-config startup-config
Confirm saving to FLASH (y/n) ? y
Copy running configuration to startup configuration
Switch is currently set to use factory default config block on next boot.
Do you want to change that to the active config block (y/n) ? y
Next boot will use active config block.
```

9. Make sure the boot configuration-block is active by using the **`show boot configuration-block`** command. If it is not active, change the boot configuration-block with the following command:

```
CN 4093(config)# boot configuration-block active
```

Secure FTP

Enterprise NOS supports Secure FTP (SFTP) to the switch. SFTP uses Secure Shell (SSH) to transfer files. SFTP encrypts both commands and data, and prevents passwords and sensitive information from being transmitted openly over the network.

All file transfer commands include SFTP support along with FTP and TFTP support. SFTP is available through the menu-based CLI, ISCLI, BBI, and SNMP.

The following examples illustrate SFTP support for ISCLI commands:

```
CN 4093# copy sftp {image1|image2|boot-image} [mgt-port|data-port]
(Copy software image from SFTP server to the switch)

CN 4093# copy sftp {ca-cert|host-cert|host-key} [mgt-port|data-port]
(Copy HTTPS certificate or host key from SFTP server to the switch)
```

Boot Strict Mode

The implementations specified in this section are compliant with National Institute of Standards and Technology (NIST) Special Publication (SP) 800-131A.

The CN4093 10 Gb Converged Scalable Switch can operate in two boot modes:

- Compatibility mode (default): This is the default switch boot mode. This mode may use algorithms and key lengths that may not be allowed/acceptable by NIST SP 800-131A specification. This mode is useful in maintaining compatibility with previous releases and in environments that have lesser data security requirements.
- Strict mode: Encryption algorithms, protocols, and key lengths in strict mode are compliant with NIST SP 800-131A specification.

When in boot strict mode, the switch uses Secure Sockets Layer (SSL)/Transport Layer Security (TLS) 1.2 protocols to ensure confidentiality of the data to and from the switch.

By default, HTTP, Telnet, and SNMPv1 and SNMPv2 are disabled on the CN4093.

Before enabling strict mode, ensure the following:

- The software version on all connected switches is Enterprise NOS 8.4.
- NIST Strict compliance is enabled on the Chassis Management Module.
- The supported protocol versions and cryptographic cipher suites between clients and servers are compatible. For example: if using SSH to connect to the switch, ensure that the SSH client supports SSHv2 and a strong cipher suite that is compliant with the NIST standard.
- Compliant Web server certificate is installed on the switch, if using BBI.
- A new self-signed certificate is generated for the switch (`CN 4093(config)# access https generate-certificate`). The new certificate is generated using 2048-bit RSA key and SHA-256 digest.
- Protocols that are not NIST SP 800-131A compliant must be disabled or not used.
- Only SSHv2 or higher is used.
- The current configuration, if any, must be saved in a location external to the switch. When the switch reboots, both the startup and running configuration are lost.

- Only protocols/algorithms compliant with NIST SP 800-131A specification are used/enabled on the switch. Please see the NIST SP 800-131A publication for details. The following table lists the acceptable protocols and algorithms:

Table 4. *Acceptable Protocols and Algorithms*

Protocol/Function	Strict Mode Algorithm	Compatibility Mode Algorithm
BGP	BGP does not comply with NIST SP 800-131A specification. When in strict mode, BGP is disabled. However, it can be enabled, if required.	Acceptable
Certificate Generation	RSA-2048 SHA-256	RSA 2048 SHA 256
Certificate Acceptance	RSA 2048 or higher SHA 224 or higher	RSA SHA, SHA2
HTTPS	TLS 1.2 only See "Acceptable Cipher Suites" on page 55;	TLS 1.0, 1.1, 1.2 See "Acceptable Cipher Suites" on page 55;
IKE		
Key Exchange	DH Group 24	DH group 1, 2, 5, 14, 24
Encryption	3DES, AES-128-CBC	3DES, AES-128-CBC
Integrity	HMAC-SHA1	HMAC-SHA1, HMAC-MD5
IPSec		
AH	HMAC-SHA1	HMAC-SHA1, HMAC-MD5
ESP	3DES, AES-128-CBC, HMAC-SHA1	3DES, AES-128-CBC, HMAC-SHA1, HMAC-MD5
LDAP	LDAP does not comply with NIST SP 800-131A specification. When in strict mode, LDAP is disabled. However, it can be enabled, if required.	Acceptable
OSPF	OSPF does not comply with NIST SP 800-131A specification. When in strict mode, OSPF is disabled. However, it can be enabled, if required.	Acceptable
RADIUS	RADIUS does not comply with NIST SP 800-131A specification. When in strict mode, RADIUS is disabled. However, it can be enabled, if required.	Acceptable
Random Number Generator	NIST SP 800-90A AES CTR DRBG	NIST SP 800-90A AES CTR DRBG
Secure NTP	Secure NTP does not comply with NIST SP 800-131A specification. When in strict mode, secure NTP is disabled. However, it can be enabled, if required.	Acceptable
SLP	SHA-256 or higher RSA/DSA 2048 or higher	

Table 4. *Acceptable Protocols and Algorithms*

Protocol/Function	Strict Mode Algorithm	Compatibility Mode Algorithm
SNMP	SNMPv3 only AES-128-CFB-128/SHA1 **Note:** Following algorithms are acceptable if you choose to support old SNMPv3 factory default users: AES-128-CFB/SHA1 DES/MD5 AES-128-CFB-128/SHA1	SNMPv1, SNMPv2, SNMPv3 DES/MD5, AES-128-CFB-128/SHA1
SSH/SFTP		
Host Key	SSH-RSA	SSH-RSA
Key Exchange	ECDH-SHA2-NISTP521 ECDH-SHA2-NISTP384 ECDH-SHA2-NISTP256 ECDH-SHA2-NISTP224 RSA2048-SHA256 DIFFIE-HELL-MAN-GROUP-EXCHANGE-SHA256 DIFFIE-HELL-MAN-GROUP-EXCHANGE-SHA1	ECDH-SHA2-NISTP521 ECDH-SHA2-NISTP384 ECDH-SHA2-NISTP256 ECDH-SHA2-NISTP224 ECDH-SHA2-NISTP192 RSA2048-SHA256 RSA1024-SHA1 DIFFIE-HELL-MAN-GROUP-EXCHANGE-SHA256 DIFFIE-HELL-MAN-GROUP-EXCHANGE-SHA1 DIFFIE-HELL-MAN-GROUP14-SHA1 DIFFIE-HELL-MAN-GROUP1-SHA1
Encryption	AES128-CTR AES128-CBC 3DES-CBC	AES128-CTR AES128-CBC RIJNDAEL128-CBC BLOWFISH-CBC 3DES-CBC ARCFOUR256 ARCFOUR128 ARCFOUR
MAC	HMAC-SHA1 HMAC-SHA1-96	HMAC-SHA1 HMAC-SHA1-96 HMAC-MD5 HMAC-MD5-96
TACACS+	TACACS+ does not comply with NIST SP 800-131A specification. When in strict mode, TACACS+ is disabled. However, it can be enabled, if required.	Acceptable

Acceptable Cipher Suites

The following cipher suites are acceptable (listed in the order of preference) when the CN4093 10 Gb Converged Scalable Switch is in compatibility mode:

Table 5. *List of Acceptable Cipher Suites in Compatibility Mode*

Cipher ID	Key Exchange	Authentication	Encryption	MAC	Cipher Name
0xC027	ECDHE	RSA	AES_128_CBC	SHA256	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256
0xC013	ECDHE	RSA	AES_128_CBC	SHA1	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA
0xC012	ECDHE	RSA	3DES	SHA1	SSL_ECDHE_RSA_WITH_3DES_EDE_CBC_SHA
0xC011	ECDHE	RSA	RC4	SHA1	SSL_ECDHE_RSA_WITH_RC4_128_SHA
0x002F	RSA	RSA	AES_128_CBC	SHA1	TLS_RSA_WITH_AES_128_CBC_SHA
0x003C	RSA	RSA	AES_128_CBC	SHA256	TLS_RSA_WITH_AES_128_CBC_SHA256
0x0005	RSA	RSA	RC4	SHA1	SSL_RSA_WITH_RC4_128_SHA
0x000A	RSA	RSA	3DES	SHA1	SSL_RSA_WITH_3DES_EDE_CBC_SHA
0x0033	DHE	RSA	AES-128_CBC	SHA1	TLS_DHE_RSA_WITH_AES_128_CBC_SHA
0x0067	DHE	RSA	AES_128_CBC	SHA256	TLS_DHE_RSA_WITH_AES_128_CBC_SHA256
0x0016	DHE	RSA	3DES	SHA1	SSL_DHE_RSA_WITH_3DES_EDE_CBC_SHA

The following cipher suites are acceptable (listed in the order of preference) when the CN4093 10 Gb Converged Scalable Switch is in strict mode:

Table 6. *List of Acceptable Cipher Suites in Strict Mode*

Cipher ID	Key Exchange	Authentication	Encryption	MAC	Cipher Name
0xC027	ECDHE	RSA	AES_128_CBC	SHA256	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256
0xC013	ECDHE	RSA	AES_128_CBC	SHA1	TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA
0xC012	ECDHE	RSA	3DES	SHA1	SSL_ECDHE_RSA_WITH_3DES_EDE_CBC_SHA
0x0033	DHE	RSA	AES-128_CBC	SHA1	TLS_DHE_RSA_WITH_AES_128_CBC_SHA
0x0067	DHE	RSA	AES_128_CBC	SHA256	TLS_DHE_RSA_WITH_AES_128_CBC_SHA256
0x0016	DHE	RSA	3DES	SHA1	SSL_DHE_RSA_WITH_3DES_EDE_CBC_SHA
0x002F	RSA	RSA	AES_128_CBC	SHA1	TLS_RSA_WITH_AES_128_CBC_SHA
0x003C	RSA	RSA	AES_128_CBC	SHA256	TLS_RSA_WITH_AES_128_CBC_SHA256
0x000A	RSA	RSA	3DES	SHA1	SSL_RSA_WITH_3DES_EDE_CBC_SHA

Configuring Strict Mode

To change the switch mode to boot strict mode, use the following command:

```
CN 4093(config)# [no] boot strict enable
```

When strict mode is enabled, you will see the following message:

```
Warning, security strict mode limits the cryptographic algorithms used by
secure protocols on this switch. Please see the documentation for full
details, and verify that peer devices support acceptable algorithms
before enabling this mode. The mode change will take effect after
reloading the switch and the configuration will be wiped during the
reload. System will enter security strict mode with default factory
configuration at next boot up.

Do you want SNMPV3 support old default users in strict mode (y/n)?
```

For SNMPv3 default users, see "SNMP Version 3" on page 561.

When strict mode is disabled, the following message is displayed:

```
Warning, disabling security strict mode. The mode change will take effect
after reloading the switch.
```

You must reboot the switch for the boot strict mode enable/disable to take effect.

Limitations

In Enterprise NOS 8.4, consider the following limitation/restrictions if you need to operate the switch in boot strict mode:

- Power ITEs and High-Availability features do not comply with NIST SP 800-131A specification.
- The CN4093 will not discover Platform agents/Common agents that are not in strict mode.
- Web browsers that do not use TLS 1.2 cannot be used.
- Limited functions of the switch managing Windows will be available.

Configuring No-Prompt Mode

If you expect to administer the switch using SNSC or another browser-based interface, you need to turn off confirmation prompts. When CLI confirmation prompts are disabled, the switch will choose the default answer. To accomplish this, use one of the following commands:

```
CN 4093(config)# [no] prompting
```

Note: This command will disable CLI confirmation prompts for current and future sessions.

```
CN 4093(config)# [no] terminal dont-ask
```

Note: This command will disable CLI confirmation prompts for the current session only. It also takes precedence over the **prompting** command - any settings configured through the **prompting** command will be disregarded for the duration of the current session.

For more details, see the *Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch Command Reference for Enterprise NOS 8.4*.

Chapter 2. Initial Setup

To help with the initial process of configuring your switch, the Enterprise NOS software includes a Setup utility. The Setup utility prompts you step-by-step to enter all the necessary information for basic configuration of the switch.

Setup can be activated manually from the command line interface any time after login: `CN 4093(config)# setup`

Information Needed for Setup

Setup requests the following information:

- Basic system information
  - Date & time
  - Whether to use Spanning Tree Group or not
- Optional configuration for each port
  - Speed, duplex, flow control, and negotiation mode (as appropriate)
  - Whether to use VLAN tagging or not (as appropriate)
- Optional configuration for each VLAN
  - Name of VLAN
  - Which ports are included in the VLAN
- Optional configuration of IP parameters
  - IP address/mask and VLAN for each IP interface
  - IP addresses for default gateway
  - Whether IP forwarding is enabled or not

Default Setup Options

The Setup prompt appears automatically whenever you login as the system administrator under the factory default settings.

1. Connect to the switch.

   After connecting, the login prompt will appear as shown here.

```
Enter login username:
Enter login password:
```

2. Enter **USERID** as the default administrator and **PASSW0RD** (with a zero) as the default password.

3. Enter the following command at the prompt:

```
CN 4093(config)# setup
```

Stopping Setup

To abort the Setup utility, press **<Ctrl+C>** during any Setup question. When you abort Setup, the system will prompt:

```
Would you like to run from top again? [y/n]
```

Enter **n** to abort Setup, or **y** to restart the Setup program at the beginning.

Restarting Setup

You can restart the Setup utility manually at any time by entering the following command at the administrator prompt:

```
CN 4093(config)# setup
```

Setup Part 1: Basic System Configuration

When Setup is started, the system prompts:

```
"Set Up" will walk you through the configuration of
  System Date and Time, Spanning Tree, Port Speed/Mode,
  VLANs, and IP interfaces. [type Ctrl-C to abort "Set Up"]
```

1. Enter **y** if you will be configuring VLANs. Otherwise enter **n**.

   If you decide not to configure VLANs during this session, you can configure them later using the configuration menus, or by restarting the Setup facility. For more information on configuring VLANs, see the *Enterprise NOS Application Guide*.

   Next, the Setup utility prompts you to input basic system information.

2. Enter the year of the current date at the prompt:

   ```
   System Date:
   Enter year [2012]:
   ```

   Enter the four-digits that represent the year. To keep the current year, press **<Enter>**.

3. Enter the month of the current system date at the prompt:

   ```
   System Date:
   Enter month [1]:
   ```

   Enter the month as a number from 1 to 12. To keep the current month, press **<Enter>**.

4. Enter the day of the current date at the prompt:

   ```
   Enter day [3]:
   ```

   Enter the date as a number from 1 to 31. To keep the current day, press **<Enter>**.

   The system displays the date and time settings:

   ```
   System clock set to 18:55:36 Wed Jan 28, 2012.
   ```

5. Enter the hour of the current system time at the prompt:

   ```
   System Time:
   Enter hour in 24-hour format [18]:
   ```

   Enter the hour as a number from 00 to 23. To keep the current hour, press **<Enter>**.

6. Enter the minute of the current time at the prompt:

   ```
   Enter minutes [55]:
   ```

   Enter the minute as a number from 00 to 59. To keep the current minute, press **<Enter>**.

7. Enter the seconds of the current time at the prompt:

```
Enter seconds [37]:
```

   Enter the seconds as a number from 00 to 59. To keep the current second, press **<Enter>**. The system then displays the date and time settings:

```
System clock set to 8:55:36 Wed Jan 28, 2012.
```

8. Turn BOOTP on or off at the prompt:

```
BootP Option:
Current BOOTP: disabled
Enter new BOOTP [d/e]:
```

   Enter **e** to enable BOOTP, or enter **d** to disable BOOTP.

9. Turn Spanning Tree Protocol on or off at the prompt:

```
Spanning Tree:
Current Spanning Tree Group 1 setting: ON
Turn Spanning Tree Group 1 OFF? [y/n]
```

   Enter **y** to turn off Spanning Tree, or enter **n** to leave Spanning Tree on.

Setup Part 2: Port Configuration

Note: When configuring port options for your switch, some prompts and options may be different.

1. Select whether you will configure VLANs and VLAN tagging for ports:

```
Port Config:
Will you configure VLANs and Tagging/Trunk-mode for ports? [y/n]
```

   If you wish to change settings for VLANs, enter **y** or enter **n** to skip VLAN configuration.

   Note: The sample screens that appear in this document might differ slightly from the screens displayed by your system. Screen content varies based on the type of chassis unit that you are using and the firmware versions and options that are installed.

2. Select the port to configure, or skip port configuration at the prompt:

   If you wish to change settings for individual ports, enter the number of the port you wish to configure. To skip port configuration, press **<Enter>** without specifying any port and go to "Setup Part 3: VLANs" on page 66.

3. Configure Gigabit Ethernet port flow parameters.

   The system prompts:

```
Gig Link Configuration:
Port Flow Control:
Current Port EXT1 flow control setting:    both
Enter new value ["rx"/"tx"/"both"/"none"]:
```

   Enter **rx** to enable receive flow control, **tx** for transmit flow control, **both** to enable both or **none** to turn flow control off for the port. To keep the current setting, press **<Enter>**.

4. Configure Gigabit Ethernet port autonegotiation mode.

   If you selected a port that has a Gigabit Ethernet connector, the system prompts:

```
Port Auto Negotiation:
Current Port EXT1 autonegotiation:         on
Enter new value ["on"/"off"]:
```

   Enter **on** to enable port autonegotiation, **off** to disable it or press **<Enter>** to keep the current setting.

5. If configuring VLANs, enable or disable VLAN tagging for the port.

   If you have selected to configure VLANs back in Part 1, the system prompts:

```
Port Tagging/Trunk-mode config (Tagged/Trunk-mode port can be a member of
multiple VLANs):
Current Tagging/Trunk-mode support: disabled
Enter new Tagging/Trunk-mode support [d/e]:
```

   Enter **d** to disable VLAN tagging for the port or enter **e** to enable VLAN tagging for the port. To keep the current setting, press **<Enter>**.

6. The system prompts you to configure the next port:

```
Enter port (INTA1-C14, EXT1-22):
```

   When you are through configuring ports, press **<Enter>** without specifying any port. Otherwise, repeat the steps in this section.

Setup Part 3: VLANs

If you chose to skip VLANs configuration back in Part 2, skip to "Setup Part 4: IP Configuration" on page 67.

1. Select the VLAN to configure, or skip VLAN configuration at the prompt:

```
VLAN Config:
Enter VLAN number from 2 to 4094, NULL at end:
```

   If you wish to change settings for individual VLANs, enter the number of the VLAN you wish to configure. To skip VLAN configuration, press **<Enter>** without typing a VLAN number and go to "Setup Part 4: IP Configuration" on page 67.

2. Enter the new VLAN name at the prompt:

```
Current VLAN name: VLAN 2
Enter new VLAN name:
```

   Entering a new VLAN name is optional. To use the pending new VLAN name, press **<Enter>**.

3. Enter the VLAN port numbers:

```
Define Ports in VLAN:
Current VLAN 2:  empty
Enter ports one per line, NULL at end:
```

   Enter each port, by port number, and confirm placement of the port into this VLAN. When you are finished adding ports to this VLAN, press **<Enter>** without specifying any port.

4. Configure Spanning Tree Group membership for the VLAN:

```
Spanning Tree Group membership:
Enter new STG index [1-128](802.1d)/[1](RSTP)/[0-32](MSTP):
```

5. The system prompts you to configure the next VLAN:

```
VLAN Config:
Enter VLAN number from 2 to 4094, NULL at end:
```

   Repeat the steps in this section until all VLANs have been configured. When all VLANs have been configured, press **<Enter>** without specifying any VLAN.

Setup Part 4: IP Configuration

The system prompts for IPv4 parameters.

Although the switch supports both IPv4 and IPv6 networks, the Setup utility permits only IPv4 configuration. For IPv6 configuration, see "Internet Protocol Version 6" on page 407.

IP Interfaces

IP interfaces are used for defining the networks to which the switch belongs.

Up to 128 IP interfaces can be configured on the CN4093 10 Gb Converged Scalable Switch (CN4093). The IP address assigned to each IP interface provides the switch with an IP presence on your network. No two IP interfaces can be on the same IP network. The interfaces can be used for connecting to the switch for remote configuration, and for routing between subnets and VLANs (if used).

Note: Interfaces 125-128 is reserved for IPv4 switch management.

1. Select the IP interface to configure, or skip interface configuration at the prompt:

```
IP Config:

IP interfaces:
Enter interface number: (1-128)
```

   If you wish to configure individual IP interfaces, enter the number of the IP interface you wish to configure. To skip IP interface configuration, press **<Enter>** without typing an interface number and go to "Default Gateways" on page 68.

2. For the specified IP interface, enter the IP address in IPv4 dotted decimal notation:

```
Current IP address:     0.0.0.0
Enter new IP address:
```

   To keep the current setting, press **<Enter>**.

3. At the prompt, enter the IPv4 subnet mask in dotted decimal notation:

```
Current subnet mask:            0.0.0.0
Enter new subnet mask:
```

   To keep the current setting, press **<Enter>**.

4. If configuring VLANs, specify a VLAN for the interface.

   This prompt appears if you selected to configure VLANs back in Part 1:

```
Current VLAN:     1
Enter new VLAN [1-4094]:
```

   Enter the number for the VLAN to which the interface belongs, or press **<Enter>** without specifying a VLAN number to accept the current setting.

5. At the prompt, enter **y** to enable the IP interface or **n** to leave it disabled:

```
Enable IP interface? [y/n]
```

6. The system prompts you to configure another interface:

```
Enter interface number: (1-128)
```

Repeat the steps in this section until all IP interfaces have been configured. When all interfaces have been configured, press **<Enter>** without specifying any interface number.

Default Gateways

1. At the prompt, select an IP default gateway for configuration or skip default gateway configuration:

```
IP default gateways:
Enter default gateway number: (1-3, 4)
```

Enter the number for the IP default gateway to be configured. To skip default gateway configuration, press **<Enter>** without typing a gateway number and go to "IP Routing" on page 69.

2. At the prompt, enter the IPv4 address for the selected default gateway:

```
Current IP address:     0.0.0.0
Enter new IP address:
```

Enter the IPv4 address in dotted decimal notation, or press **<Enter>** without specifying an address to accept the current setting.

3. At the prompt enter **y** to enable the default gateway or **n** to leave it disabled:

```
Enable default gateway? [y/n]
```

4. The system prompts you to configure another default gateway:

```
Enter default gateway number: (1-4)
```

Repeat the steps in this section until all default gateways have been configured. When all default gateways have been configured, press **<Enter>** without specifying any number.

IP Routing

When IP interfaces are configured for the various IP subnets attached to your switch, IP routing between them can be performed entirely within the switch. This eliminates the need to send inter-subnet communication to an external router device. Routing on more complex networks, where subnets may not have a direct presence on the CN4093, can be accomplished through configuring static routes or by letting the switch learn routes dynamically.

This part of the Setup program prompts you to configure the various routing parameters.

At the prompt, enable or disable forwarding for IP Routing:

```
Enable IP forwarding? [y/n]
```

Enter **y** to enable IP forwarding. To disable IP forwarding, enter **n**. To keep the current setting, press **<Enter>**.

Setup Part 5: Final Steps

1. When prompted, decide whether to restart Setup or continue:

```
Would you like to run from top again? [y/n]
```

   Enter **y** to restart the Setup utility from the beginning or **n** to continue.

2. When prompted, decide whether you wish to review the configuration changes:

```
Review the changes made? [y/n]
```

   Enter **y** to review the changes made during this session of the Setup utility. Enter **n** to continue without reviewing the changes. We recommend that you review the changes.

3. Next, decide whether to apply the changes at the prompt:

```
Apply the changes? [y/n]
```

   Enter **y** to apply the changes or **n** to continue without applying. Changes are normally applied.

4. At the prompt, decide whether to make the changes permanent:

```
Save changes to flash? [y/n]
```

   Enter **y** to save the changes to flash. Enter **n** to continue without saving the changes. Changes are normally saved at this point.

5. If you do not apply or save the changes, the system prompts whether to abort them:

```
Abort all changes? [y/n]
```

   Enter **y** to discard the changes. Enter **n** to return to the "`Apply the changes?`" prompt.

   Note: After initial configuration is complete, it is recommended that you change the default passwords as shown in "Changing the Switch Passwords" on page 90.

Optional Setup for Telnet Support

Note: This step is optional. Perform this procedure only if you are planning on connecting to the CN4093 through a remote Telnet connection.

1. Telnet is enabled by default. To change the setting, use the following command:

```
CN 4093(config)# no access telnet
```

Chapter 3. Switch Software Management

The switch software image is the executable code running on the CN4093. A version of the image comes pre-installed on the device. As new versions of the image are released, you can upgrade the software running on your switch. To get the latest version of software supported for your CN4093, go to the following website:

http://www.ibm.com/support/

To determine the software version currently used on the switch, use the following switch command:

```
CN 4093# show boot
```

The typical upgrade process for the software image consists of the following steps:

- Load a new software image and boot image onto an FTP, SFTP or TFTP server on your network.
- Transfer the new images to your switch.
- Specify the new software image as the one which will be loaded into switch memory the next time a switch reset occurs.
- Reset the switch.

For instructions on the typical upgrade process using the ENOS ISCLI or BBI, see "Loading New Software to Your Switch" on page 74.

CAUTION:
Although the typical upgrade process is all that is necessary in most cases, upgrading from (or reverting to) some versions of Lenovo Enterprise Network Operating System requires special steps prior to or after the software installation process. Please be sure to follow all applicable instructions in the release notes document for the specific software release to ensure that your switch continues to operate as expected after installing new software.

Loading New Software to Your Switch

The CN4093 can store up to two different switch software images (called `image1` and `image2`) as well as special boot software (called `boot`). When you load new software, you must specify where it is placed: either into `image1`, `image2` or `boot`.

For example, if your active image is currently loaded into `image1`, you would probably load the new image software into `image2`. This lets you test the new software and reload the original active image (stored in `image1`), if needed.

CAUTION:
When you upgrade the switch software image, always load the new boot image and the new software image before you reset the switch. If you do not load a new boot image, your switch might not boot properly (To recover, see "Recover from a Failed Image Upgrade using TFTP" on page 81 **or** "Recovering from a Failed Image Upgrade using XModem Download" on page 83**).**

To load a new software image to your switch, you will need the following:

- The image and boot software loaded on an FTP, SFTP or TFTP server on your network.

  Note: Be sure to download both the new boot file and the new image file.

- The hostname or IP address of the FTP, SFTP or TFTP server

  Note: The DNS parameters must be configured if specifying hostnames.

- The name of the new software image or boot file

When the software requirements are met, use one of the following procedures to download the new software to your switch. You can use the ISCLI or the BBI to download and activate new software.

Loading Software via the ISCLI

1. Since FTP is disabled by default, you need to enable it. In Privileged EXEC mode, enter the following command:

```
CN 4093# access ftp enable
```

2. Copy the image to your switch, specifying the method for loading the software (FTP, SFTP, or TFTP) and the CN4093 destination (`image1`, `image2`, or `boot-image`) by entering the following command:

```
CN 4093# copy {tftp|ftp|sftp} {image1|image2|boot-image}
```

3. Enter the hostname or IP address of the FTP, SFTP or TFTP server.

```
Address or name of remote host: <name or IP address>
```

4. Enter the name of the new software file on the server.

```
Source file name: <filename>
```

The exact form of the name will vary by server. However, the file location is normally relative to the FTP, SFTP or TFTP directory (for example, `tftpboot`).

5. If required by the FTP, SFTP or TFTP server, enter the appropriate username and password.

6. The switch will prompt you to confirm your request.

   Once confirmed, the software will begin loading into the switch.

7. When loading is complete, use the following commands to enter Global Configuration mode to select which software image (`image1` or `image2`) you want to run in switch memory for the next reboot:

```
Router# configure terminal
Router(config)# boot image {image1|image2}
```

   The system will then verify which image is set to be loaded at the next reset:

```
Next boot will use switch software image1 instead of image2.
```

8. Reboot the switch to run the new software:

```
Router(config)# reload
```

   The system prompts you to confirm your request. Once confirmed, the switch will reboot to use the new software.

Loading Software via BBI

You can use the Browser-Based Interface to load software onto the CN4093. The software image to load can reside in one of the following locations:

- FTP server
- TFTP server
- SFTP server
- Local computer

After you log onto the BBI, perform the following steps to load a software image:

1. Click the Configure context tab in the toolbar.
2. In the Navigation Window, select System > Config/Image Control.

   The Switch Image and Configuration Management page appears.

3. If you are loading software from your computer (HTTP client), skip this step and go to the next. Otherwise, if you are loading software from an FTP, SFTP, or TFTP server, enter the server's information in the FTP, SFTP, or TFTP Settings section.
4. In the Image Settings section, select the image version you want to replace (Image for Transfer).
   - If you are loading software from an FTP, SFTP, or TFTP server, enter the file name and click **Get Image**.
   - If you are loading software from your computer, click **Browse**.
   - In the File Upload Dialog, select the file and click **OK**. Then click **Download via Browser**.

   Once the image has loaded, the page refreshes to show the new software.

Updating Software on vLAG Switches

When updating the software and boot images for switches configured with vLAG, first:

- Make sure that the spanning tree root switch is not one of the vLAG switches
- Shut down of ports should done under the port configuration
- Follow the shut down order of the ports

  a. ISL links

  b. vLAG links

  c. vLAG health check (MGT port)

Then follow this procedure to update the software on vLAG switches:

1. On Switch 2 (the original Secondary switch), shut down all links ISL, vLAG links, and vLAG HC. This is equivalent to powering off Switch 2.
   - All the traffic will failover to Switch 1 (the original Primary switch.).
   - After the shutdown of links on Switch 2, there will be N-S traffic loss of around ~0.16 seconds.
2. Upgrade Switch 2 with the new image. Use FTP, STFP, or TFTP to copy the new ENOS and boot images onto the switch. For more details, see "Loading New Software to Your Switch" on page 74.
   - After Switch 2 comes up, vLAG HC will be up and vLAG mismatch will happen with vLAG ports down (since it is still Secondary).
   - The traffic will still be forwarding via Switch 1 (the original Primary switch).
3. On Switch 1 (the original Primary switch), shut down all links ISL, vLAG links, and vLAG HC. This is equivalent to powering off Switch 1 (the original Primary switch)
   - All the traffic will failover to Switch 2, which will assume the vLAG operation role of Primary.
   - After all the links are up on Switch 2, there will be N-S traffic loss of around ~70 seconds due to spanning trees reconverging.
4. Upgrade Switch 1 (the original Primary switch with the new ENOS image. Use FTP, STFP, or TFTP to copy the new ENOS and boot images onto the switch. For more details, see "Loading New Software to Your Switch" on page 74.
   - After Switch 1 comes up, vLAG HC, ISL, and vLAG links will be up, and Switch 1 will assume the vLAG operation role of Secondary.
   - All the traffic will now follow the hash and load balance settings between Switch 1 and Switch 2.
   - There will be N-S traffic loss of around ~0.05 seconds.
5. Change the operational role of the vLAG switches to match the final topology by reloading Switch 2.
   - There will be N-S traffic loss of around ~0.102 seconds.

- Switch 1 will reassume the vLAG Primary role and Switch 2 will reassume the vLAG Secondary role.

6. Make sure that Switch 1 is now the vLAG primary switch and Switch 2 is now the vLAG secondary switch using the following command:

```
CN 4093> show vlag information
```

The Boot Management Menu

The Boot Management menu allows you to switch the software image, reset the switch to factory defaults, or to recover from a failed software download.

You can interrupt the boot process and enter the Boot Management menu from the serial console port. When the system displays Memory Test, press **<Shift+B>**. The Boot Management menu appears.

```
Boot Management Menu
        I - Change booting image
        C - Change configuration block
        R - Boot in recovery mode (tftp and xmodem download of images to
recover switch)
        Q - Reboot
        E - Exit
Please choose your menu option:
```

The Boot Management menu allows you to perform the following actions:

The Boot Management menu allows you to perform the following actions:

- To change the booting image, press **I** and follow the screen prompts.
- To change the configuration block, press **C** and follow the screen prompts.
- To boot in recovery mode press **R**. For more details see "Boot Recovery Mode" on page 80.
- To restart the boot process from the beginning, press **Q**.
- To exit the Boot Management menu, press **E**. The booting process continues.

Boot Recovery Mode

The Boot Recovery Mode allows you to recover from a failed software or boot image upgrade using TFTP or XModem download.

To enter Boot Recovery Mode you must select "Boot in recovery mode" option from the Boot Management Menu by pressing **R**.

```
Entering Rescue Mode.
Please select one of the following options:
        T) Configure networking and tftp download an image
        X) Use xmodem 1K to serial download an image
        P) Physical presence (low security mode)
        R) Reboot
        E) Exit

Option? :
```

The Boot Recovery Mode menu allows you to perform the following actions:

- To recover from a failed software or boot image upgrade using TFTP, press **T** and follow the screen prompts. For more details, see "Recover from a Failed Image Upgrade using TFTP" on page 81.
- To recover from a failed software or boot image upgrade using XModem download, press **X** and follow the screen prompts. For more details, see "Recovering from a Failed Image Upgrade using XModem Download" on page 83.
- To enable the loading of an unofficial image, press **P** and follow the screen prompts. For more details, see "Physical Presence" on page 85.
- To restart the boot process from the beginning, press **R**.
- To exit Boot Recovery Mode menu, press **E**. The boot process continues.

Recover from a Failed Image Upgrade using TFTP

Use the following procedure to recover from a failed image upgrade using TFTP:

1. Connect a PC to the console port of the switch.
2. Open a terminal emulator program that supports Telnet protocol (for example, HyperTerminal, SecureCRT or PuTTY) and input the proper host name (IP address) and port to connect to the console port of the switch.
3. Boot the switch and access the Boot Management menu by pressing **<Shift+B>** while the Memory Test is in progress and the dots are being displayed.
4. Enter Boot Recovery Mode by pressing **R**. The Recovery Mode menu will appear.
5. To start the recovery process using TFTP, press **T**. The following message will appear:

```
Performing TFTP rescue. Please answer the following questions (enter 'q'
to quit):
```

6. Enter the IP address of the management port:

```
IP addr :
```

7. Enter the network mask of the management port:

```
Netmask :
```

8. Enter the gateway of the management port:

```
Gateway :
```

9. Enter the IP address of the TFTP server:

```
Server addr:
```

10. Enter the filename of the image:

```
Image Filename:
```

11. If the file is a software image, enter an image number:

```
Install image as image 1 or 2 (hit return to just boot image):
```

After the procedure is complete, the Recovery Mode menu will be re-displayed.

Below is an example of a successful recovery procedure using TFTP:

```
Entering Rescue Mode.
Please select one of the following options:
        T) Configure networking and tftp download an image
        X) Use xmodem 1K to serial download an image
        P) Physical presence (low security mode)
        R) Reboot
        E) Exit

Option? : t
Performing TFTP rescue. Please answer the following questions (enter 'q'
to quit):
IP addr :10.241.6.4
Netmask :255.255.255.128
Gateway :10.241.6.66
Server addr:10.72.97.135
Image Filename: CN4093-8.3.1.0_OS.img
        Netmask : 255.255.255.128
        Gateway : 10.241.6.66
Configuring management port.....................
Installing image CN4093-8.3.1.0_OS.img from TFTP server 10.72.97.135

Extracting images ... Do *NOT* power cycle the switch.
Installing Application: Image signature verified.
Install image as image 1 or 2 (hit return to just boot image): 2
Installing image as image2: 100%

Image2 updated succeeded
Updating install log. File CN4093-8.3.1.0_OS.img installed from
10.72.97.135 at 15:29:30 on 12-3-2015
Please select one of the following options:
        T) Configure networking and tftp download an image
        X) Use xmodem 1K to serial download an image
        P) Physical presence (low security mode)
        R) Reboot
        E) Exit

Option? :
```

Recovering from a Failed Image Upgrade using XModem Download

Use the following procedure to recover from a failed image upgrade.

1. Connect a PC to the serial port of the switch.
2. Open a terminal emulator program that supports Xmodem download (for example, HyperTerminal, SecureCRT or PuTTY) and select the following serial port characteristics:
   - Speed: 9600 bps
   - Data Bits: 8
   - Stop Bits: 1
   - Parity: None
   - Flow Control: None
3. Boot the switch and access the Boot Management menu by pressing **<Shift+B>** while the Memory Test is in progress and the dots are being displayed.
4. Enter Boot Recovery Mode by pressing **R**. The Recovery Mode menu will appear.
5. Press **X** for Xmodem download. You will see the following display:

```
Running xmodem rescue.....
```

6. When you see the following message, change the Serial Port speed to 115200 bps:

```
Change the baud rate to 115200 bps and hit the <ENTER> key before
initiating the download.
```

7. Press **<Enter>** to set the system into download accept mode. When the readiness meter displays (a series of "C" characters), start Xmodem on your terminal emulator. You will see a display similar to the following:

```
... Waiting for the <Enter> key to be hit before the download can start...
CC
```

8. Select the image to download. Xmodem initiates the file transfer. When download is complete, you are asked to change the Serial Port speed back to 9600 bps:

```
Change the baud rate back to 9600 bps, hit the <ENTER> key
```

9. Press **<Enter>** to start installing the image. If the file is a software image, enter the image number:

```
Install image as image 1 or 2 (hit return to just boot image):
```

The image install will begin. After the procedure is complete, the Recovery Mode menu will be re-displayed.

```
Extracting images ... Do *NOT* power cycle the switch.
Installing Root Filesystem:
Image signature verified.  100%
Installing Kernel:
Image signature verified.  100%
Installing Device Tree:
Image signature verified.  100%
Installing Boot Loader:  100%
Updating install log. File image installed from xmodem at 18:06:02 on
13-3-2015
Please select one of the following options:
        T) Configure networking and tftp download an image
        X) Use xmodem 1K to serial download an image
        P) Physical presence (low security mode)
        R) Reboot
        E) Exit

Option? :
```

Boot image recovery is complete.

Physical Presence

Use the following procedure to enable the installation of unofficial images on the switch:

1. Connect a PC to the console port of the switch.
2. Open a terminal emulator program that supports Telnet protocol (for example, HyperTerminal, SecureCRT or PuTTY) and input the proper host name (IP address) and port to connect to the console port of the switch.
3. Boot the switch and access the Boot Management menu by pressing **<Shift+B>** while the Memory Test is in progress and the dots are being displayed.
4. Enter Boot Recovery Mode by pressing **R**. The Recovery Mode menu will appear.
5. To begin the Physical Presence procedure, press **P**. The following warning message will appear:

```
WARNING: the following test is used to determine physical presence and if
completed will put the switch in low security mode.
```

6. You will be prompted for confirmation:

```
Do you wish to continue y/n?
```

7. A security test will be performed. The system location (blue) LED will blink a number of times between 1 and 12. Enter that number:

```
Hit a key to start the test. The blue location LED will blink a number of
times.

...........
How many times did the LED blink?
```

8. After entering the correct number, the Recovery Mode menu will re-appear. To install an unofficial image use one of the following procedures:
   - TFTP (for details, see page 81)
   - XModem Download (for details, see page 83)

   Note: You have three attempts to successfully complete the security test. After three incorrect attempts, the switch will reboot.

   Note: After the test is completed, the switch will be put in low security mode. This mode will allow you to install unofficial images on the switch. To revert to normal security mode, you must reboot the switch or press **P** again in the Recovery Mode menu.

Part 2: Securing the Switch

Chapter 4. Securing Administration

This chapter discusses different methods of securing local and remote administration on the CN4093 10 Gb Converged Scalable Switch (CN4093):

Changing the Switch Passwords

It is recommended that you change the administrator and user passwords after initial configuration and as regularly as required under your network security policies.

To change the administrator password, you must login using the administrator password.

Note: If you download user and password information to a switch running a version of ENOS earlier than 8.4, or if you revert the switch to a version of ENOS earlier than 8.4, your passwords will not be transferred because the encryption algorithm changed.

Changing the Default Administrator Password

The administrator has complete access to all menus, information, and configuration commands, including the ability to change both the user and administrator passwords.

The default administrator account is `USERID`. The default password for the administrator account is `PASSW0RD` (with a zero). To change the administrator password, use the following procedure:

1. Connect to the switch and log in as the administrator.
2. Use the following command to change the administrator password:

```
CN 4093(config)# access user administrator-password <password>
```

Changing the Default User Password

The user login has limited control of the switch. Through a user account, you can view switch information and statistics, but you can't make configuration changes.

The default password for the user account is `user`. This password can be changed from the user account. The administrator can change all passwords, as shown in the following procedure.

1. Connect to the switch and log in as the administrator.
2. Use the following command to change the user password:

```
CN 4093(config)# access user user-password <password>
```

Secure Shell and Secure Copy

Because using Telnet does not provide a secure connection for managing a CN4093, Secure Shell (SSH) and Secure Copy (SCP) features have been included for CN4093 management. SSH and SCP use secure tunnels to encrypt and secure messages between a remote administrator and the switch.

SSH is a protocol that enables remote administrators to log securely into the CN4093 over a network to execute management commands.

SCP is typically used to copy files securely from one machine to another. SCP uses SSH for encryption of data on the network. On a CN4093, SCP is used to download and upload the switch configuration via secure channels.

Although SSH and SCP are disabled by default, enabling and using these features provides the following benefits:

- Identifying the administrator using Name/Password
- Authentication of remote administrators
- Authorization of remote administrators
- Determining the permitted actions and customizing service for individual administrators
- Encryption of management messages
- Encrypting messages between the remote administrator and switch
- Secure copy support

The Enterprise NOS implementation of SSH supports both versions 1.5 and 2.0 and supports SSH clients version 1.5 - 2.x. The following SSH clients have been tested:

- SSH 1.2.23 and SSH 1.2.27 for Linux (freeware)
- SecureCRT 3.0.2 and SecureCRT 3.0.3 for Windows NT (Van Dyke Technologies, Inc.)
- F-Secure SSH 1.1 for Windows (Data Fellows)
- Putty SSH
- Cygwin OpenSSH
- Mac X OpenSSH
- Solaris 8 OpenSSH
- AxeSSH SSHPro
- SSH Communications Vandyke SSH A
- F-Secure

Configuring SSH/SCP Features on the Switch

SSH and SCP are disabled by default. To change the setting, using the following procedures.

Note: To use SCP, you must first enable SSH.

To Enable or Disable the SSH Feature

Begin a Telnet session from the console port and enter the following commands:

```
CN 4093(config)# ssh enable          (Turn SSH on)

CN 4093(config)# no ssh enable       (Turn SSH off)
```

To Enable or Disable SCP

Enter the following command to enable or disable SCP:

```
CN 4093(config)# [no] ssh scp-enable
```

Configuring the SCP Administrator Password

To configure the SCP-only administrator password, enter the following command (the default password is `admin`):

```
CN 4093(config)# [no] ssh scp-password
Changing SCP-only Administrator password; validation required...
Enter current administrator password: <password>
Enter new SCP-only administrator password: <new password>
Re-enter new SCP-only administrator password: <new password>
New SCP-only administrator password accepted.
```

Using SSH and SCP Client Commands

This section shows the format for using some common client commands.

To Log In to the Switch from the Client

Syntax:

```
>> ssh [-4|-6] <switch IP address>
   -or-
>> ssh [-4|-6] <login name>@<switch IP address>
```

Note: The `-4` option (the default) specifies that an IPv4 switch address will be used. The `-6` option specifies IPv6.

Example**:**

```
>> ssh scpadmin@205.178.15.157
```

To Copy the Switch Configuration File to the SCP Host

Syntax**:**

```
>> scp [-4|-6] <username>@<switch IP address>:getcfg <local filename>
```

Example:

```
>> scp scpadmin@205.178.15.157:getcfg ad4.cfg
```

To Load a Switch Configuration File from the SCP Host

Syntax:

```
>> scp [-4|-6] <local filename> <username>@<switch IP address>:putcfg
```

Example:

```
>> scp ad4.cfg scpadmin@205.178.15.157:putcfg
```

To Apply and Save the Configuration

When loading a configuration file to the switch, the `apply` and `save` commands are still required, in order for the configuration commands to take effect. The `apply` and `save` commands may be entered manually on the switch, or by using SCP commands.

Syntax:

```
>> scp [-4|-6] <local filename> <username>@<switch IP address>:putcfg_apply
>> scp [-4|-6] <local filename> <username>@<switch IP address>:putcfg_apply_save
```

Example:

```
>> scp ad4.cfg scpadmin@205.178.15.157:putcfg_apply
>> scp ad4.cfg scpadmin@205.178.15.157:putcfg_apply_save
```

- The CLI `diff` command is automatically executed at the end of `putcfg` to notify the remote client of the difference between the new and the current configurations.
- `putcfg_apply` runs the `apply` command after the `putcfg` is done.
- `putcfg_apply_save` saves the new configuration to the flash after `putcfg_apply` is done.
- The `putcfg_apply` and `putcfg_apply_save` commands are provided because extra `apply` and `save` commands are usually required after a `putcfg`; however, an SCP session is not in an interactive mode.

To Copy the Switch Image and Boot Files to the SCP Host

Syntax**:**

```
>> scp [-4|-6] <username>@<switch IP address>:getimg1 <local filename>
>> scp [-4|-6] <username>@<switch IP address>:getimg2 <local filename>
>> scp [-4|-6] <username>@<switch IP address>:getboot <local filename>
```

Example:

```
>> scp scpadmin@205.178.15.157:getimg1 6.1.0_os.img
```

To Load Switch Configuration Files from the SCP Host

Syntax:

```
>> scp [-4|-6] <local filename>  <username>@<switch IP address>:putimg1
>> scp [-4|-6] <local filename>  <username>@<switch IP address>:putimg2
>> scp [-4|-6] <local filename>  <username>@<switch IP address>:putboot
```

Example:

```
>> scp 6.1.0_os.img scpadmin@205.178.15.157:putimg1
```

SSH and SCP Encryption of Management Messages

The following encryption and authentication methods are supported for SSH and SCP:

- Server Host Authentication: Client RSA authenticates the switch at the beginning of every connection
- Key Exchange: RSA
- Encryption: 3DES-CBC, DES
- User Authentication: Local password authentication, RADIUS

Generating an RSA Host Key for SSH Access

To support the SSH server feature, an RSA host key is required. The host key is 2048 bits and is used to identify the CN4093.

When the SSH server is first enabled and applied, the switch automatically generates the RSA host key and stores it in FLASH memory.

To configure an RSA host key, first connect to the CN4093 through the console port (commands are not available via external Telnet connection), and enter the following command to generate it manually.

```
CN 4093(config)# ssh generate-host-key      (Generates the host key)
```

When the switch reboots, it will retrieve the host key from the FLASH memory.

Notes:

- The switch will perform only one session of key/cipher generation at a time. Thus, an SSH/SCP client will not be able to log in if the switch is performing key generation at that time. Also, key generation will fail if an SSH/SCP client is logging in at that time.
- Because the switch software only generates RSA keys, if there is already a DSA-based SSH key on the switch, this key will remain on the switch and not be replaced until you run the `ssh generate-host key` command to generate an RSA key.

SSH/SCP Integration with RADIUS Authentication

SSH/SCP is integrated with RADIUS authentication. After the RADIUS server is enabled on the switch, all subsequent SSH authentication requests will be redirected to the specified RADIUS servers for authentication. The redirection is transparent to the SSH clients.

SSH/SCP Integration with TACACS+ Authentication

SSH/SCP is integrated with TACACS+ authentication. After the TACACS+ server is enabled on the switch, all subsequent SSH authentication requests will be redirected to the specified TACACS+ servers for authentication. The redirection is transparent to the SSH clients.

End User Access Control

Enterprise NOS allows an administrator to define end user accounts that permit end users to perform operation tasks via the switch CLI commands. Once end user accounts are configured and enabled, the switch requires username/password authentication.

For example, an administrator can assign a user, who can then log into the switch and perform operational commands (effective only until the next switch reboot).

Considerations for Configuring End User Accounts

- A maximum of 20 user IDs are supported on the switch.
- Enterprise NOS supports end user support for Console, Telnet, BBI, and SSHv1/v2 access to the switch.
- If RADIUS authentication is used, the user password on the Radius server will override the user password on the CN4093. Also note that the password change command modifies only the user switch password on the switch and has no effect on the user password on the Radius server. Radius authentication and user password cannot be used concurrently to access the switch.
- Passwords can be up to 64 characters in length for Telnet, SSH, Console, and Web access.

Strong Passwords

The administrator can require use of Strong Passwords for users to access the CN4093. Strong Passwords enhance security because they make password guessing more difficult.

The following rules apply when Strong Passwords are enabled:

- Minimum length: 8 characters; maximum length: 64 characters
- Must contain at least one uppercase alphabet
- Must contain at least one lowercase alphabet
- Must contain at least one number
- Must contain at least one special character:

  Supported special characters: ! " # % & ' ( ) ; < = >> ? [\ ] * + , - . / : ^ _ { | } ~
- Cannot be same as the username

When strong password is enabled, users can still access the switch using the old password but will be advised to change to a strong password while attempting to log in.

The strong password requirement can be enabled using the following command:

```
CN 4093(config)# access user strong-password enable
```

The administrator can choose the number of days allowed before each password expires. When a strong password expires, the user is allowed to log in one last time (last time) to change the password. A warning provides advance notice for users to change the password.

User Access Control Menu

The end-user access control commands allow you to configure end-user accounts.

Setting Up User IDs

Up to 20 user IDs can be configured in the User ID menu.

```
CN 4093(config)# access user 1 name <1-8 characters>
CN 4093(config)# access user 1 password

Changing user1 password; validation required:
Enter current admin password: <current administrator password>
Enter new user1 password: <new user password>
Re-enter new user1 password: <new user password>
New user1 password accepted.
```

Defining a User's Access Level

The end user is by default assigned to the user access level (also known as class of service, or CoS). CoS for all user accounts have global access to all resources except for User CoS, which has access to view only resources that the user owns. For more information, see Table 7 on page 106.

To change the user's level, enter the class of service `cos` command:

```
CN 4093(config)# access user 1 level {user|operator|administrator}
```

Validating a User's Configuration

```
CN 4093# show access user uid 1
```

Enabling or Disabling a User

An end user account must be enabled before the switch recognizes and permits login under the account. Once enabled, the switch requires any user to enter both username and password.

```
CN 4093(config)# [no] access user 1 enable
```

Locking Accounts

To protect the switch from unauthorized access, the account lockout feature can be enabled. By default, account lockout is disabled. To enable this feature, ensure the strong password feature is enabled (*See* "Strong Passwords" on page 96). Then use the following command:

```
CN 4093(config)# access user strong-password lockout
```

After multiple failed login attempts, the switch locks the user account if lockout has been enabled on the switch.

Re-enabling Locked Accounts

The administrator can re-enable a locked account by reloading the switch or by using the following command:

```
CN 4093(config)# access user strong-password clear local user lockout
username <user name>
```

However, the above command cannot be used to re-enable an account disabled by the administrator.

To re-enable all locked accounts, use the following command:

```
CN 4093(config)# access user strong-password clear local user lockout all
```

Listing Current Users

The **show access user** command displays defined user accounts and whether or not each user is currently logged into the switch.

```
CN 4093# show access user

Usernames:
  user     - Enabled - offline
  oper     - Disabled - offline
  admin    - Always Enabled - online 1 session

Current User ID table:
  1: name USERID  , ena, cos admin   , password valid, offline
  2: name jane    , ena, cos user    , password valid, online
  3: name john    , ena, cos user    , password valid, online
```

Logging In to an End User Account

Once an end user account is configured and enabled, the user can login to the switch, using the username/password combination. The level of switch access is determined by the Class of Service established for the end user account.

Protected Mode

Protected Mode settings allow the switch administrator to block the management module from making configuration changes that affect switch operation. The switch retains control over those functions.

The following management module functions are disabled when Protected Mode is turned on:

- External Ports: Enabled/Disabled
- External management over all ports: Enabled/Disabled
- Restore Factory Defaults
- New Static IP Configuration

In this release, configuration of the functions listed above are restricted to the local switch when you turn Protected Mode on. In future releases, individual control over each function may be added.

Note: Before you turn Protected Mode on, make sure that external management (Telnet) access to one of the switch's IP interfaces is enabled.

Use the following command to turn Protected Mode on:

```
CN 4093(config)# protected-mode enable
```

If you lose access to the switch through the external ports, use the console port to connect directly to the switch, and configure an IP interface with Telnet access.

Stacking Mode

When the switch is in stacking mode, Protected Mode is automatically enabled for three of the four Protected Mode functions, and the following module functions are disabled:

- External Ports (Enabled)
- External management over all ports (Enabled)
- Restore Factory Defaults

Stack members and stack Master can get their IP addresses from the advanced management module (AMM). Stack can be managed using external ports or using the AMM management port.

If required, the functionality of new static IP configuration can also be disabled by turning off Protected Mode (`CN 4093(config)# **no protected-mode enable**`) and turning it back on (`CN 4093(config)# **protected-mode enable**`).

Maintenance Mode

There are times when Lenovo support needs to access your switch in maintenance mode. To enable this, enter the command:

```
CN 4093(config)# maint-internal
```

When prompted, enter the admin password.

The Lenovo support person will then enter the maintenance mode password.

This introduces a second level of administration authorization before the Lenovo support representative enters the maintenance mode password, making the switch more secure and available for enhanced debugging.

Chapter 5. Authentication & Authorization Protocols

Secure switch management is needed for environments that perform significant management functions across the Internet. The following are some of the functions for secured IPv4 management and device access:

Note: Enterprise NOS 8.4 does not support IPv6 for RADIUS, TACACS+, or LDAP.

RADIUS Authentication and Authorization

Enterprise NOS supports the RADIUS (Remote Authentication Dial-in User Service) method to authenticate and authorize remote administrators for managing the switch. This method is based on a client/server model. The Remote Access Server (RAS)—the switch—is a client to the back-end database server. A remote user (the remote administrator) interacts only with the RAS, not the back-end server and database.

RADIUS authentication consists of the following components:

- A protocol with a frame format that utilizes UDP over IP (based on RFC 2138 and 2866)
- A centralized server that stores all the user authorization information
- A client, in this case, the switch

The CN4093—acting as the RADIUS client—communicates to the RADIUS server to authenticate and authorize a remote administrator using the protocol definitions specified in RFC 2138 and 2866. Transactions between the client and the RADIUS server are authenticated using a shared key that is not sent over the network. In addition, the remote administrator passwords are sent encrypted between the RADIUS client (the switch) and the back-end RADIUS server.

How RADIUS Authentication Works

1. Remote administrator connects to the switch and provides user name and password.
2. Using Authentication/Authorization protocol, the switch sends request to authentication server.
3. Authentication server checks the request against the user ID database.
4. Using RADIUS protocol, the authentication server instructs the switch to grant or deny administrative access.

Configuring RADIUS on the Switch

Use the following procedure to configure Radius authentication on your CN4093.

1. Turn RADIUS authentication on, then configure the Primary and Secondary RADIUS servers.

```
CN 4093(config)# radius-server primary-host 10.10.1.1
CN 4093(config)# radius-server secondary-host 10.10.1.2
```

2. Configure the RADIUS secret.

```
CN 4093(config)# radius-server primary-host 10.10.1.1 key
<1-32 character secret>
CN 4093(config)# radius-server secondary-host 10.10.1.2 key
<1-32 character secret>
CN 4093(config)# radius-server enable
```

CAUTION:
If you configure the RADIUS secret using any method other than through the console port, the secret may be transmitted over the network as clear text.

3. If desired, you may change the default UDP port number used to listen to RADIUS.

   The well-known port for RADIUS is 1645.

```
CN 4093(config)# radius-server port <UDP port number>
```

4. Configure the number retry attempts for contacting the RADIUS server, and the timeout period.

```
CN 4093(config)# radius-server retransmit 3
CN 4093(config)# radius-server timeout 5
```

RADIUS Authentication Features in Enterprise NOS

Enterprise NOS supports the following RADIUS authentication features:

- Supports RADIUS client on the switch, based on the protocol definitions in RFC 2138 and RFC 2866.
- Allows a RADIUS secret password of up to 32 characters.
- Supports *secondary authentication server* so that when the primary authentication server is unreachable, the switch can send client authentication requests to the secondary authentication server. Use the following command to show the currently active RADIUS authentication server:

```
CN 4093# show radius-server
```

- Supports user-configurable RADIUS server retry and time-out values:
  - Time-out value = 1-10 seconds
  - Retries = 1-3

  The switch will time out if it does not receive a response from the RADIUS server within 1-10 seconds. The switch automatically retries connecting to the RADIUS server 1-3 times before it declares the server down.
- Supports user-configurable RADIUS application port. The default is UDP port 1645. UDP port 1812, based on RFC 2138, is also supported.
- Allows network administrator to define privileges for one or more specific users to access the switch at the RADIUS user database.

Switch User Accounts

The user accounts listed in Table 7 can be defined in the RADIUS server dictionary file.

Table 7. *User Access Levels*

User Account	Description and Tasks Performed	Password
User	The User has no direct responsibility for switch management. He/she can view all switch status information and statistics but cannot make any configuration changes to the switch.	`user`
Operator	In addition to User capabilities, the Operator has limited switch management access, including the ability to make temporary, operational configuration changes to some switch features, and to reset switch ports (other than management ports).	`oper`
Administrator (`USERID`)	The super-user Administrator has complete access to all menus, information, and configuration commands on the switch, including the ability to change both the user and administrator passwords.	`PASSW0RD`

RADIUS Attributes for Enterprise NOS User Privileges

When the user logs in, the switch authenticates his/her level of access by sending the RADIUS access request, that is, the client authentication request, to the RADIUS authentication server.

If the remote user is successfully authenticated by the authentication server, the switch will verify the *privileges* of the remote user and authorize the appropriate access. The administrator has two options: to allow *local* access via Telnet, SSH, HTTP, or HTTPS; to allow *secure local* access via console, Telnet, SSH, or BBI. Secure local access provides access to the switch when the RADIUS servers cannot be reached.

The default CN4093 setting for local access and secure local access is `disabled`. Local access is always enabled on the console port.

Whether local access is enabled or not, you can always access the switch via the console port by using `noradius` as the RADIUS username. You can then enter the username and password configured on the switch. If you are trying to connect via SSH/Telnet/HTTP/HTTPS, there are two possibilities:

- Local access is enabled: The switch acts like it is connecting via console.
- Secure local access is enabled: You must enter the username: `noradius`. The switch checks if RADIUS server is reachable. If it is reachable, then you must authenticate via remote authentication server. Only if RADIUS server is not reachable, you will be prompted for local user/password to be authenticated against these local credentials.

All user privileges, other than those assigned to the Administrator, have to be defined in the RADIUS dictionary. RADIUS attribute 6 which is built into all RADIUS servers defines the administrator. The file name of the dictionary is RADIUS vendor-dependent. The following RADIUS attributes are defined for Enterprise NOS user privileges levels:

Table 8. *Enterprise NOS-proprietary Attributes for RADIUS*

User Name/Access	User-Service-Type	Value
User	*Vendor-supplied*	255
Operator	*Vendor-supplied*	252
Administrator (`USERID`)	*Vendor-supplied*	6

TACACS+ Authentication

Enterprise NOS supports authentication, authorization, and accounting with networks using the Cisco Systems TACACS+ protocol. The CN4093 functions as the Network Access Server (NAS) by interacting with the remote client and initiating authentication and authorization sessions with the TACACS+ access server. The remote user is defined as someone requiring management access to the CN4093 either through a data or management port.

TACACS+ offers the following advantages over RADIUS:

- TACACS+ uses TCP-based connection-oriented transport; whereas RADIUS is UDP-based. TCP offers a connection-oriented transport, while UDP offers best-effort delivery. RADIUS requires additional programmable variables such as re-transmit attempts and time-outs to compensate for best-effort transport, but it lacks the level of built-in support that a TCP transport offers.
- TACACS+ offers full packet encryption whereas RADIUS offers password-only encryption in authentication requests.
- TACACS+ separates authentication, authorization and accounting.

How TACACS+ Authentication Works

TACACS+ works much in the same way as RADIUS authentication as described on page 104.

1. Remote administrator connects to the switch and provides user name and password.
2. Using Authentication/Authorization protocol, the switch sends request to authentication server.
3. Authentication server checks the request against the user ID database.
4. Using TACACS+ protocol, the authentication server instructs the switch to grant or deny administrative access.

   During a session, if additional authorization checking is needed, the switch checks with a TACACS+ server to determine if the user is granted permission to use a particular command.f

TACACS+ Authentication Features in Enterprise NOS

Authentication is the action of determining the identity of a user, and is generally done when the user first attempts to log in to a device or gain access to its services. Enterprise NOS supports ASCII inbound login to the device. PAP, CHAP and ARAP login methods, TACACS+ change password requests, and one-time password authentication are not supported.

Authorization

Authorization is the action of determining a user's privileges on the device, and usually takes place after authentication.

The default mapping between TACACS+ authorization levels and Enterprise NOS management access levels is shown in Table 9. The authorization levels listed in this table must be defined on the TACACS+ server.

Table 9. *Default TACACS+ Authorization Levels*

Enterprise NOS User Access Level	TACACS+ Level
`user`	0
`oper`	3
`admin (USERID)`	6

Alternate mapping between TACACS+ authorization levels and Enterprise NOS management access levels is shown in Table 10. Use the following command to use the alternate TACACS+ authorization levels:

```
CN 4093(config)# tacacs-server privilege-mapping
```

Table 10. *Alternate TACACS+ Authorization Levels*

Enterprise NOS User Access Level	TACACS+ Level
`user`	0–1
`oper`	6–8
`admin (USERID)`	14–15

You can customize the mapping between TACACS+ privilege levels and CN4093 management access levels. Use the following command to manually map each TACACS+ privilege level (0-15) to a corresponding CN4093 management access level:

```
CN 4093(config)# tacacs-server user-mapping
```

If the remote user is successfully authenticated by the authentication server, the switch verifies the *privileges* of the remote user and authorizes the appropriate access.

Local Access

The administrator has an option to allow *local* access via Telnet using the command:

```
CN 4093(config)# tacacs-server backdoor
```

The default value for Telnet access is `disabled`. The administrator also can enable *secure local access* to allow access if both the primary and the secondary TACACS+ servers fail to respond. The command for this is:

```
CN 4093(config)# tacacs-server secure-backdoor
```

Note: To obtain the TACACS+ local access password for your switch, contact your Service and Support line.

Accounting

Accounting is the action of recording a user's activities on the device for the purposes of billing and/or security. It follows the authentication and authorization actions. If the authentication and authorization is not performed via TACACS+, there are no TACACS+ accounting messages sent out.

You can use TACACS+ to record and track software login access, configuration changes, and interactive commands.

The CN4093 supports the following TACACS+ accounting attributes:

- protocol (console/telnet/ssh/http)
- start_time
- stop_time
- elapsed_time
- disc-cause

Note: When using the Browser-Based Interface, the TACACS+ Accounting Stop records are sent only if the **Quit** button on the browser is clicked.

Command Authorization and Logging

When TACACS+ Command Authorization is enabled, ENOS configuration commands are sent to the TACACS+ server for authorization. Use the following command to enable TACACS+ Command Authorization:

```
CN 4093(config)# tacacs-server command-authorization
```

When TACACS+ Command Logging is enabled, ENOS configuration commands are logged on the TACACS+ server. Use the following command to enable TACACS+ Command Logging:

```
CN 4093(config)# tacacs-server command-logging
```

The following examples illustrate the format of Enterprise NOS commands sent to the TACACS+ server:

```
authorization request, cmd=cfgtree, cmd-arg=/cfg/l3/if
accounting request, cmd=/cfg/l3/if, cmd-arg=1
authorization request, cmd=cfgtree, cmd-arg=/cfg/l3/if/ena
accounting request, cmd=/cfg/l3/if/ena
authorization request, cmd=cfgtree, cmd-arg=/cfg/l3/if/addr
accounting request, cmd=/cfg/l3/if/addr, cmd-arg=10.90.90.91

authorization request, cmd=apply
accounting request, cmd=apply
```

The following rules apply to TACACS+ command authorization and logging:

- Only commands from a Console, Telnet, or SSH connection are sent for authorization and logging. SNMP, BBI, or file-copy commands (for example, TFTP or sync) are not sent.
- Only leaf-level commands are sent for authorization and logging. For example:

  ```
  CN 4093(config)#
  ```

  is not sent, but the following command is sent:

  ```
  CN 4093(config)# tacacs-server command-logging
  ```

- The full path of each command is sent for authorization and logging. For example:

  ```
  CN 4093(config)# tacacs-server command-logging
  ```

- Command arguments are not sent for authorization.
- Only executed commands are logged.
- Invalid commands are checked by Enterprise NOS and are not sent for authorization or logging.

- Authorization is performed on each leaf-level command separately. If the user issues multiple commands at once, each command is sent separately as a full path.
- Only the following global commands are sent for authorization and logging:
  - `diff`
  - `ping`
  - `revert`
  - `telnet`
  - `traceroute`

TACACS+ Password Change

Enterprise NOS supports TACACS+ password change. When enabled, users can change their passwords after successful TACACS+ authorization. Use the following command to enable or disable this feature:

```
CN 4093(config)# [no] tacacs-server password-change
```

Use the following commands to change the password for the primary and secondary TACACS+ servers:

```
CN 4093(config)# tacacs-server chpassp     (Change primary TACACS+ password)
CN 4093(config)# tacacs-server chpasss     (Change secondary TACACS+ password)
```

Configuring TACACS+ Authentication on the Switch

1. Configure the IPv4 addresses of the Primary and Secondary TACACS+ servers, and enable TACACS authentication.

```
Enter primary server IPv4 address:
CN 4093(config)# tacacs-server primary-host 10.10.1.1
CN 4093(config)# tacacs-server primary-host mgt-port

Enter secondary server IPv4 address:
CN 4093(config)# tacacs-server secondary-host 10.10.1.1
CN 4093(config)# tacacs-server secondary-host data-port

CN 4093(config)# tacacs-server enable
```

2. Configure the TACACS+ secret and second secret.

```
CN 4093(config)# tacacs-server primary-host 10.10.1.1 key
<1-32 character secret>
CN 4093(config)# tacacs-server secondary-host 10.10.1.2 key
<1-32 character secret>
```

If you configure the TACACS+ secret using any method other than a direct console connection, the secret may be transmitted over the network as clear text.

3. If desired, you may change the default TCP port number used to listen to TACACS+. The well-known port for TACACS+ is 49.

```
CN 4093(config)# tacacs-server port <TCP port number>
```

4. Configure the number of retry attempts, and the timeout period.

```
CN 4093(config)# tacacs-server retransmit 3
CN 4093(config)# tacacs-server timeout 5
```

5. Configure custom privilege-level mapping (optional).

```
CN 4093(config)# tacacs-server user-mapping 2 user
CN 4093(config)# tacacs-server user-mapping 3 user
CN 4093(config)# tacacs-server user-mapping 4 user
CN 4093(config)# tacacs-server user-mapping 5 oper
```

LDAP Authentication and Authorization

Enterprise NOS supports the LDAP (Lightweight Directory Access Protocol) method to authenticate and authorize remote administrators to manage the switch. LDAP is based on a client/server model. The switch acts as a client to the LDAP server. A remote user (the remote administrator) interacts only with the switch, not the back-end server and database.

LDAP authentication consists of the following components:

- A protocol with a frame format that utilizes TCP over IP
- A centralized server that stores all the user authorization information
- A client, in this case, the switch

Each entry in the LDAP server is referenced by its Distinguished Name (DN). The DN consists of the user-account name concatenated with the LDAP domain name. If the user-account name is John, the following is an example DN:

```
uid=John,ou=people,dc=domain,dc=com
```

Configuring the LDAP Server

CN4093 user groups and user accounts must reside within the same domain. On the LDAP server, configure the domain to include CN4093 user groups and user accounts, as follows:

- User Accounts:

  Use the *uid* attribute to define each individual user account.

- User Groups:

  Use the *members* attribute in the *groupOfNames* object class to create the user groups. The first word of the common name for each user group must be equal to the user group names defined in the CN4093, as follows:

  - `admin (USERID)`
  - `oper`
  - `user`

Configuring LDAP Authentication on the Switch

1. Turn LDAP authentication on, then configure the Primary and Secondary LDAP servers.

```
CN 4093(config)# ldap-server enable
CN 4093(config)# ldap-server primary-host 10.10.1.1
CN 4093(config)# ldap-server secondary-host 10.10.1.2
```

2. Configure the domain name.

```
CN 4093(config)# ldap-server domain <ou=people,dc=my-domain,dc=com>
```

3. If desired, you may change the default TCP port number used to listen to LDAP.

   The well-known port for LDAP is 389.

```
CN 4093(config)# ldap-server port <1-65000>
```

4. Configure the number of retry attempts for contacting the LDAP server and the timeout period.

```
CN 4093(config)# ldap-server retransmit 3 (number of server retries)
CN 4093(config)# ldap-server timeout 10 (enter the timeout period in seconds)
```

5. You may change the default LDAP attribute (uid) or add a custom attribute. For instance, Microsoft's Active Directory requires the `cn` (common name) attribute.

```
CN 4093(config)# ldap-server attribute username <1-128 alpha-numeric characters>
```

Chapter 6. 802.1X Port-Based Network Access Control

Port-Based Network Access control provides a means of authenticating and authorizing devices attached to a LAN port that has point-to-point connection characteristics. It prevents access to ports that fail authentication and authorization. This feature provides security to ports of the CN4093 10 Gb Converged Scalable Switch (CN4093) that connect to blade servers.

The following topics are discussed in this section:

Extensible Authentication Protocol over LAN

Enterprise NOS can provide user-level security for its ports using the IEEE 802.1X protocol, which is a more secure alternative to other methods of port-based network access control. Any device attached to an 802.1X-enabled port that fails authentication is prevented access to the network and denied services offered through that port.

The 802.1X standard describes port-based network access control using Extensible Authentication Protocol over LAN (EAPoL). EAPoL provides a means of authenticating and authorizing devices attached to a LAN port that has point-to-point connection characteristics and of preventing access to that port in cases of authentication and authorization failures.

EAPoL is a client-server protocol that has the following components:

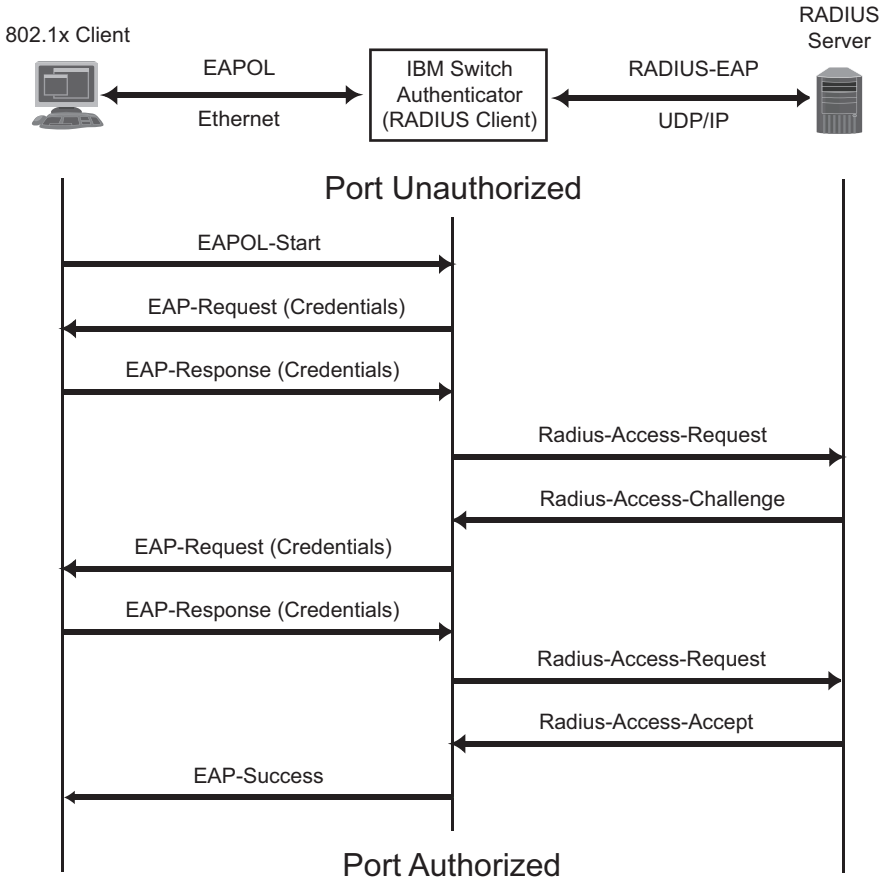
- Supplicant or Client
  The Supplicant is a device that requests network access and provides the required credentials (user name and password) to the Authenticator and the Authenticator Server.
- Authenticator
  The Authenticator enforces authentication and controls access to the network. The Authenticator grants network access based on the information provided by the Supplicant and the response from the Authentication Server. The Authenticator acts as an intermediary between the Supplicant and the Authentication Server: requesting identity information from the client, forwarding that information to the Authentication Server for validation, relaying the server's responses to the client, and authorizing network access based on the results of the authentication exchange. The CN4093 acts as an Authenticator.
- Authentication Server
  The Authentication Server validates the credentials provided by the Supplicant to determine if the Authenticator should grant access to the network. The Authentication Server may be co-located with the Authenticator. The CN4093 relies on external RADIUS servers for authentication.

Upon a successful authentication of the client by the server, the 802.1X-controlled port transitions from unauthorized to authorized state, and the client is allowed full access to services through the port. When the client sends an EAP-Logoff message to the authenticator, the port will transition from authorized to unauthorized state.

EAPoL Authentication Process

The clients and authenticators communicate using Extensible Authentication Protocol (EAP), which was originally designed to run over PPP, and for which the IEEE 802.1X Standard has defined an encapsulation method over Ethernet frames, called EAP over LAN (EAPOL). Figure 1 shows a typical message exchange initiated by the client.

Figure 1. Authenticating a Port Using EAPoL



EAPoL Message Exchange

During authentication, EAPOL messages are exchanged between the client and the CN4093 authenticator, while RADIUS-EAP messages are exchanged between the CN4093 authenticator and the RADIUS server.

Authentication is initiated by one of the following methods:

- The CN4093 authenticator sends an EAP-Request/Identity packet to the client
- The client sends an EAPOL-Start frame to the CN4093 authenticator, which responds with an EAP-Request/Identity frame.

The client confirms its identity by sending an EAP-Response/Identity frame to the CN4093 authenticator, which forwards the frame encapsulated in a RADIUS packet to the server.

The RADIUS authentication server chooses an EAP-supported authentication algorithm to verify the client's identity, and sends an EAP-Request packet to the client via the CN4093 authenticator. The client then replies to the RADIUS server with an EAP-Response containing its credentials.

Upon a successful authentication of the client by the server, the 802.1X-controlled port transitions from unauthorized to authorized state, and the client is allowed full access to services through the controlled port. When the client later sends an EAPOL-Logoff message to the CN4093 authenticator, the port transitions from authorized to unauthorized state.

If a client that does not support 802.1X connects to an 802.1X-controlled port, the CN4093 authenticator requests the client's identity when it detects a change in the operational state of the port. The client does not respond to the request, and the port remains in the unauthorized state.

Note: When an 802.1X-enabled client connects to a port that is not 802.1X-controlled, the client initiates the authentication process by sending an EAPOL-Start frame. When no response is received, the client retransmits the request for a fixed number of times. If no response is received, the client assumes the port is in authorized state, and begins sending frames, even if the port is unauthorized.

EAPoL Port States

The state of the port determines whether the client is granted access to the network, as follows:

- **Unauthorized**
  While in this state the port discards all ingress and egress traffic except EAP packets.
- **Authorized**
  When the client is successfully authenticated, the port transitions to the authorized state allowing all traffic to and from the client to flow normally.
- **Force Unauthorized**
  You can configure this state that denies all access to the port.
- **Force Authorized**
  You can configure this state that allows full access to the port.

Guest VLAN

The guest VLAN provides limited access to unauthenticated ports. The guest VLAN can be configured using the following command:

```
CN 4093(config)# dot1x guest-vlan ?
```

Client ports that have not received an EAPOL response are placed into the Guest VLAN, if one is configured on the switch. Once the port is authenticated, it is moved from the Guest VLAN to its configured VLAN.

When Guest VLAN enabled, the following considerations apply while a port is in the unauthenticated state:

- The port is placed in the guest VLAN.
- The Port VLAN ID (PVID) is changed to the Guest VLAN ID.
- Port tagging is disabled on the port.

Supported RADIUS Attributes

The 802.1X Authenticator relies on external RADIUS servers for authentication with EAP. Table 11lists the RADIUS attributes that are supported as part of RADIUS-EAP authentication based on the guidelines specified in Annex D of the 802.1X standard and RFC 3580.

Table 11. *Support for RADIUS Attributes*

#	Attribute	Attribute Value	A-R	A-A	A-C	A-R
1	User-Name	The value of the Type-Data field from the supplicant's EAP-Response/Identity message. If the Identity is unknown (i.e. Type-Data field is zero bytes in length), this attribute will have the same value as the Calling-Station-Id.	1	0-1	0	0
4	NAS-IP-Address	IPv4 address of the authenticator used for Radius communication.	1	0	0	0
5	NAS-Port	Port number of the authenticator port to which the supplicant is attached.	1	0	0	0
24	State	Server-specific value. This is sent unmodified back to the server in an Access-Request that is in response to an Access-Challenge.	0-1	0-1	0-1	0
30	Called-Station-ID	The MAC address of the authenticator encoded as an ASCII string in canonical format, such as 000D5622E3 9F.	1	0	0	0
31	Calling-Station-ID	The MAC address of the supplicant encoded as an ASCII string in canonical format, such as 00034B436206.	1	0	0	0
64	Tunnel-Type	Only VLAN (type 13) is currently supported (for 802.1X RADIUS VLAN assignment). The attribute must be untagged (the Tag field must be 0).	0	0-1	0	0
65	Tunnel-Medium-Type	Only 802 (type 6) is currently supported (for 802.1X RADIUS VLAN assignment). The attribute must be untagged (the Tag field must be 0).	0	0-1	0	0

Table 11. *Support for RADIUS Attributes (continued)*

#	Attribute	Attribute Value	A-R	A-A	A-C	A-R
81	Tunnel-Private-Group-ID	VLAN ID (1-4094). When 802.1X RADIUS VLAN assignment is enabled on a port, if the RADIUS server includes the tunnel attributes defined in RFC 2868 in the Access-Accept packet, the switch will automatically place the authenticated port in the specified VLAN. Reserved VLANs (such as for management) may not be specified. The attribute must be untagged (the Tag field must be 0).	0	0-1	0	0
79	EAP-Message	Encapsulated EAP packets from the supplicant to the authentication server (Radius) and vice-versa. The authenticator relays the decoded packet to both devices.	1+	1+	1+	1+
80	Message-Authenticator	Always present whenever an EAP-Message attribute is also included. Used to integrity-protect a packet.	1	1	1	1
87	NAS-Port-ID	Name assigned to the authenticator port, e.g. Server1_Port3	1	0	0	0

Legend: RADIUS Packet Types: A-R (Access-Request), A-A (Access-Accept), A-C (Access-Challenge), A-R (Access-Reject)

RADIUS Attribute Support:

- 0 This attribute MUST NOT be present in a packet.
- 0+ Zero or more instances of this attribute MAY be present in a packet.
- 0-1 Zero or one instance of this attribute MAY be present in a packet.
- 1 Exactly one instance of this attribute MUST be present in a packet.
- 1+ One or more of these attributes MUST be present.

EAPoL Configuration Guidelines

When configuring EAPoL, consider the following guidelines:

- The 802.1X port-based authentication is currently supported only in point-to-point configurations, that is, with a single supplicant connected to an 802.1X-enabled switch port.
- When 802.1X is enabled, a port has to be in the authorized state before any other Layer 2 feature can be operationally enabled. For example, the STG state of a port is operationally disabled while the port is in the unauthorized state.
- The 802.1X supplicant capability is not supported. Therefore, none of its ports can successfully connect to an 802.1X-enabled port of another device, such as another switch, that acts as an authenticator, unless access control on the remote port is disabled or is configured in forced-authorized mode. For example, if a CN4093 is connected to another CN4093, and if 802.1X is enabled on both switches, the two connected ports must be configured in force-authorized mode.
- Unsupported 802.1X attributes include Service-Type, Session-Timeout, and Termination-Action.
- RADIUS accounting service for 802.1X-authenticated devices or users is not currently supported.
- Configuration changes performed using SNMP and the standard 802.1X MIB will take effect immediately.

Chapter 7. Access Control Lists

Access Control Lists (ACLs) are filters that permit or deny traffic for security purposes. They can also be used with QoS to classify and segment traffic in order to provide different levels of service to different traffic types. Each filter defines the conditions that must match for inclusion in the filter, and also the actions that are performed when a match is made.

Enterprise NOS 8.4 supports the following ACLs:

- IPv4 ACLs

  Up to 640 ACLs are supported for networks that use IPv4 addressing. IPv4 ACLs are configured using the following CLI menu:

```
CN 4093(config)# access-control list <IPv4 ACL number>
```

- IPv6 ACLs

  Up to 128 ACLs are supported for networks that use IPv6 addressing. IPv6 ACLs are configured using the following CLI menu:

```
CN 4093(config)# access-control list6 <IPv6 ACL number>
```

- Management ACLs

  Up to 128 MACLs are supported. ACLs for the different types of management protocols (Telnet, HTTPS, etc.) provide greater granularity for securing management traffic.

  Management ACLs are configured using the following command:

```
CN 4093(config)# access-control macl <MACL number>
```

- VLAN Maps (VMaps)

  Up to 128 VLAN Maps are supported for attaching filters to VLANs rather than ports. See "VLAN Maps" on page 136 for details.

```
CN 4093(config)# access-control vmap <vmap number>
```

Summary of Packet Classifiers

ACLs allow you to classify packets according to a variety of content in the packet header (such as the source address, destination address, source port number, destination port number, and others). Once classified, packet flows can be identified for more processing.

Regular ACLs, and VMaps allow you to classify packets based on the following packet attributes:

- Ethernet header options (for regular ACLs and VMaps only)
  - Source MAC address
  - Destination MAC address
  - VLAN number and mask
  - Ethernet type (ARP, IPv4, MPLS, RARP, etc.)
  - Ethernet Priority (the IEEE 802.1p Priority)
- IPv4 header options (for regular ACLs and VMaps only)
  - Source IPv4 address and subnet mask
  - Destination IPv4 address and subnet mask
  - Type of Service value
  - IP protocol number or name as shown in Table 12:

Table 12. *Well-Known Protocol Types*

Number	Protocol Name
1	icmp
2	igmp
6	tcp
17	udp
89	ospf
112	vrrp

- TCP/UDP header options (for all ACLs)
  - TCP/UDP application source port as shown in Table 13.

Table 13. *Well-Known Application Ports*

Port	TCP/UDP Application	Port	TCP/UDP Application	Port	TCP/UDP Application
20	ftp-data	79	finger	179	bgp
21	ftp	80	http	194	irc
22	ssh	109	pop2	220	imap3
23	telnet	110	pop3	389	ldap
25	smtp	111	sunrpc	443	https
37	time	119	nntp	520	rip
42	name	123	ntp	554	rtsp
43	whois	143	imap	1645/1812	Radius
53	domain	144	news	1813	Radius Accounting
69	tftp	161	snmp	1985	hsrp
70	gopher	162	snmptrap		

  - TCP/UDP application destination port and mask as shown in Table 13.
  - TCP/UDP flag value as shown in Table 14.

Table 14. *Well-Known TCP flag values*

Flag	Value
URG	0x0020
ACK	0x0010
PSH	0x0008
RST	0x0004
SYN	0x0002
FIN	0x0001

- Packet format (for regular ACLs and VMaps only)
  - Ethernet format (eth2, SNAP, LLC)
  - Ethernet tagging format
  - IP format (IPv4)
- Egress port packets (for all ACLs)

Summary of ACL Actions

Once classified using ACLs, the identified packet flows can be processed differently. For each ACL, an *action* can be assigned. The action determines how the switch treats packets that match the classifiers assigned to the ACL. CN4093 ACL actions include the following:

- Pass or Drop the packet
- Re-mark the packet with a new DiffServ Code Point (DSCP)
- Re-mark the 802.1p field
- Set the COS queue

Assigning Individual ACLs to a Port

Once you configure an ACL, you must assign the ACL to the appropriate ports. Each port can accept multiple ACLs, and each ACL can be applied for multiple ports. ACLs can be assigned individually, or in groups.

To assign an individual ACL to a port, use the following IP interface commands:

```
CN 4093(config)# interface port <port>
CN 4093(config-if)# access-control list <IPv4 ACL number>
CN 4093(config-ip)# access-control list6 <IPv6 ACL number>
```

When multiple ACLs are assigned to a port, higher-priority ACLs are considered first, and their action takes precedence over lower-priority ACLs. ACL order of precedence is discussed in the next section.

To create and assign ACLs in groups, see "ACL Groups" on page 129.

ACL Order of Precedence

When multiple ACLs are assigned to a port, they are evaluated in numeric sequence, based on the ACL number. Lower-numbered ACLs take precedence over higher-numbered ACLs. For example, ACL 1 (if assigned to the port) is evaluated first and has top priority.

If multiple ACLs match the port traffic, only the action of the one with the lowest ACL number is applied. The others are ignored.

The ACL number is the sole factor in determining ACL order of precedence. The order in which ACLs are applied to a port does not affect the order of precedence, nor does the ACL Group number (see "ACL Groups" on page 129), the order in which an ACL is assigned to an ACL Group, or the order in which the ACL Group is assigned to a port.

If no assigned ACL matches the port traffic, no ACL action is applied.

ACL Groups

To assist in organizing multiple ACLs and assigning them to ports, you can place ACLs into ACL Groups, thereby defining complex traffic profiles. ACLs and ACL Groups can then be assigned on a per-port basis. Any specific ACL can be assigned to multiple ACL Groups, and any ACL or ACL Group can be assigned to multiple ports. If, as part of multiple ACL Groups, a specific ACL is assigned to a port multiple times, only one instance is used. The redundant entries are ignored.

- **Individual ACLs**
  The CN4093 supports up to 256 ACLs. Each ACL defines one filter rule for matching traffic criteria. Each filter rule can also include an action (permit or deny the packet). For example:

ACL 1: VLAN = 1 SIP = 10.10.10.1 (255.255.255.0) Action = permit

- **Access Control List Groups**
  An Access Control List Group (ACL Group) is a collection of ACLs. For example:

ACL Group 1
ACL 1: VLAN = 1 SIP = 10.10.10.1 (255.255.255.0) Action = permit
ACL 2: VLAN = 2 SIP = 10.10.10.2 (255.255.255.0) Action = deny
ACL 3: Priority = 7 DIP = 10.10.10.3 (255.255.255.0) Action = permit

  ACL Groups organize ACLs into traffic profiles that can be more easily assigned to ports. The CN4093 supports up to 256 ACL Groups.

Note: ACL Groups are used for convenience in assigning multiple ACLs to ports. ACL Groups have no effect on the order in which ACLs are applied (see "ACL Order of Precedence" on page 128). All ACLs assigned to the port (whether individually assigned or part of an ACL Group) are considered as individual ACLs for the purposes of determining their order of precedence.

Assigning ACL Groups to a Port

To assign an ACL Group to a port, use the following commands:

```
CN 4093(config)# interface port <port number>
CN 4093(config-if)# access-control group <ACL group number>
CN 4093(config-if)# exit
```

ACL Metering and Re-Marking

You can define a profile for the aggregate traffic flowing through the switch by configuring a QoS meter (if desired) and assigning ACLs to ports.

Note: When you add ACLs to a port, make sure they are ordered correctly in terms of precedence (see "ACL Order of Precedence" on page 128).

Actions taken by an ACL are called *In-Profile* actions. You can configure additional In-Profile and Out-of-Profile actions on a port. Data traffic can be metered, and re-marked to ensure that the traffic flow provides certain levels of service in terms of bandwidth for different types of network traffic.

Metering

QoS metering provides different levels of service to data streams through user-configurable parameters. A meter is used to measure the traffic stream against a traffic profile which you create. Thus, creating meters yields In-Profile and Out-of-Profile traffic for each ACL, as follows:

- **In-Profile**–If there is no meter configured or if the packet conforms to the meter, the packet is classified as In-Profile.
- **Out-of-Profile**–If a meter is configured and the packet does not conform to the meter (exceeds the committed rate or maximum burst rate of the meter), the packet is classified as Out-of-Profile.

Using meters, you set a Committed Rate in Kbps (1000 bits per second in each Kbps). All traffic within this Committed Rate is In-Profile. Additionally, you can set a Maximum Burst Size that specifies an allowed data burst larger than the Committed Rate for a brief period. These parameters define the In-Profile traffic.

Meters keep the sorted packets within certain parameters. You can configure a meter on an ACL, and perform actions on metered traffic, such as packet re-marking.

Re-Marking

Re-marking allows for the treatment of packets to be reset based on new network specifications or desired levels of service. You can configure the ACL to re-mark a packet as follows:

- Change the DSCP value of a packet, used to specify the service level that traffic should receive.
- Change the 802.1p priority of a packet.

ACL Port Mirroring

For regular ACLs and VMaps, packets that match an ACL on a specific port can be mirrored to another switch port for network diagnosis and monitoring.

The source port for the mirrored packets cannot be a portchannel, but may be a member of a portchannel.

The destination port to which packets are mirrored must be a physical port.

If the ACL or VMap has an action (permit, drop etc.) assigned, it cannot be used to mirror packets for that ACL.

Use the following commands to add mirroring to an ACL:

- For regular ACLs:

```
CN 4093(config)# access-control list <ACL number> mirror port <destination port>
```

  The ACL must be also assigned to it target ports as usual (see "Assigning Individual ACLs to a Port" on page 128, or "Assigning ACL Groups to a Port" on page 129).

- For VMaps (see "VLAN Maps" on page 136):

```
CN 4093(config)# access-control vmap <VMap number> mirror port <monitor destination port>
```

Viewing ACL Statistics

ACL statistics display how many packets have "hit" (matched) each ACL. Use ACL statistics to check filter performance or to debug the ACL filter configuration.

You must enable statistics for each ACL that you wish to monitor:

```
CN 4093(config)# access-control list <ACL number> statistics
```

ACL Logging

ACLs are generally used to enhance port security. Traffic that matches the characteristics (source addresses, destination addresses, packet type, etc.) specified by the ACLs on specific ports is subject to the actions (chiefly permit or deny) defined by those ACLs. Although switch statistics show the number of times particular ACLs are matched, the ACL logging feature can provide additional insight into actual traffic patterns on the switch, providing packet details in the system log for network debugging or security purposes.

Enabling ACL Logging

By default, ACL logging is disabled. Enable or disable ACL logging on a per-ACL basis as follows:

```
CN 4093(config)# [no] access-control list <IPv4 ACL number> log
CN 4093(config)# [no] access-control list6 <IPv6 ACL number> log
```

Logged Information

When ACL logging is enabled on any particular ACL, the switch will collect information about packets that match the ACL. The information collected depends on the ACL type:

- For IP-based ACLs, information is collected regarding
  - Source IP address
  - Destination IP address
  - TCP/UDP port number
  - ACL action
  - Number of packets logged

  For example:

  ```
  Sep 27  4:20:28 DUT3 NOTICE  ACL-LOG: %IP ACCESS LOG: list
  ACL-IP-12-IN denied tcp 1.1.1.1 (0) -> 200.0.1.2 (0), 150
  packets.
  ```

- For MAC-based ACLs, information is collected regarding
  - Source MAC address
  - Source IP address
  - Destination IP address
  - TCP/UDP port number
  - ACL action
  - Number of packets logged

  For example:

  ```
  Sep 27  4:25:38 DUT3 NOTICE  ACL-LOG: %MAC ACCESS LOG: list
  ACL-MAC-12-IN permitted tcp 1.1.1.2 (0) (12,
  00:ff:d7:66:74:62) -> 200.0.1.2 (0) (00:18:73:ee:a7:c6), 32
  packets.
  ```

Rate Limiting Behavior

Because ACL logging can be CPU-intensive, logging is rate-limited. By default, the switch will log only 10 matching packets per second. This pool is shared by all log-enabled ACLs. The global rate limit can be changed as follows:

```
CN 4093(config)# access-control log rate-limit <1-1000>
```

Where the limit is specified in packets per second.

Log Interval

For each log-enabled ACL, the first packet that matches the ACL initiates an immediate message in the system log. Beyond that, additional matches are subject to the log interval. By default, the switch will buffer ACL log messages for a period of 300 seconds. At the end of that interval, all messages in the buffer are written to the system log. The global interval value can be changed as follows:

```
CN 4093(config)# access-control log interval <5-600>
```

Where the interval rate is specified in seconds.

In any given interval, packets that have identical log information are condensed into a single message. However, the packet count shown in the ACL log message represents only the logged messages, which due to rate-limiting, may be significantly less than the number of packets actually matched by the ACL.

Also, the switch is limited to 64 different ACL log messages in any interval. Once the threshold is reached, the oldest message will be discarded in favor of the new message, and an overflow message will be added to the system log.

ACL Logging Limitations

ACL logging reserves packet queue 1 for internal use. Features that allow remapping packet queues (such as CoPP) may not behave as expected if other packet flows are reconfigured to use queue 1.

ACL Configuration Examples

ACL Example 1

Use this configuration to block traffic to a specific host. All traffic that ingresses on port EXT1 is denied if it is destined for the host at IP address `100.10.1.1`.

1. Configure an Access Control List.

```
CN 4093(config)# access-control list 1 ipv4 destination-ip-address
100.10.1.1
CN 4093(config)# access-control list 1 action deny
```

2. Add ACL 1 to port EXT1.

```
CN 4093(config)# interface port EXT1
CN 4093(config-if)# access-control list 1
CN 4093(config-if)# exit
```

ACL Example 2

Use this configuration to block traffic from a network destined for a specific host address. All traffic that ingresses in port EXT2 with source IP from class 100.10.1.0/24 and destination IP 200.20.2.2 is denied.

1. Configure an Access Control List.

```
CN 4093(config)# access-control list 2 ipv4 source-ip-address 100.10.1.0
    255.255.255.0
CN 4093(config)# access-control list 2 ipv4 destination-ip-address
    200.20.2.2 255.255.255.255
CN 4093(config)# access-control list 2 action deny
```

2. Add ACL 2 to port EXT2.

```
CN 4093(config)# interface port EXT2
CN 4093(config-if)# access-control list 2
CN 4093(config-if)# exit
```

ACL Example 3

This configuration blocks traffic from a network that is destined for a specific egress port. All traffic that ingresses port EXT1 from the network 100.10.1.0/24 and is destined for port 3 is denied.

1. Configure an Access Control List.

```
CN 4093(config)# access-control list 4 ipv4 source-ip-address 100.10.1.0
    255.255.255.0
CN 4093(config)# access-control list 4 egress-port 3
CN 4093(config)# access-control list 4 action deny
```

2. Add ACL 4 to port EXT1.

```
CN 4093(config)# interface port EXT1
CN 4093(config-if)# access-control list 4
CN 4093(config-if)# exit
```

VLAN Maps

A VLAN map (VMAP) is an ACL that can be assigned to a VLAN or VM group rather than to a switch port as with regular ACLs. This is particularly useful in a virtualized environment where traffic filtering and metering policies must follow virtual machines (VMs) as they migrate between hypervisors.

VMAPs are configured using the following ISCLI command path:

```
CN 4093(config)# access-control vmap <VMAP ID (1-128)>

  action         Set filter action
  egress-port    Set to filter for packets egressing this port
  ethernet       Ethernet header options
  ipv4           IP version 4 header options
  meter          ACL metering configuration
  mirror         Mirror options
  packet-format  Set to filter specific packet format types
  re-mark        ACL re-mark configuration
  statistics     Enable access control list statistics
  tcp-udp        TCP and UDP filtering options
```

The CN4093 supports up to 128 VMAPs.

Individual VMAP filters are configured in the same fashion as regular ACLs, except that VLANs cannot be specified as a filtering criteria (unnecessary, since the VMAP are assigned to a specific VLAN or associated with a VM group VLAN).

Once a VMAP filter is created, it can be assigned or removed using the following configuration commands:

- For a regular VLAN:

```
CN 4093(config)# vlan <VLAN ID>
CN 4093(config-vlan)# [no] vmap <VMap ID> [intports|extports]
```

- For a VM group (see "VM Group Types" on page 284):

```
CN 4093(config)# [no] virt vmgroup <ID> vmap <VMap ID>
[intports|extports]
```

Note: Each VMAP can be assigned to only one VLAN or VM group. However, each VLAN or VM group may have multiple VMAPs assigned to it.

When the optional `intports` or `extports` parameter is specified, the action to add or remove the vMAP is applies for either the internal downlink ports or external uplink ports only. If omitted, the operation will be applied to all ports in the associated VLAN or VM group.

Note: VMAPs have a lower priority than port-based ACLs. If both an ACL and a VMAP match a particular packet, both filter actions will be applied as long as there is no conflict. In the event of a conflict, the port ACL will take priority, though switch statistics will count matches for both the ACL and VMAP.

VMap Example

In this example, Ethernet type 2 traffic from VLAN 3 server ports is mirrored to a network monitor on port 4.

```
CN 4093(config)# access-control vmap 21 packet-format ethernet
ethernet-type2
CN 4093(config)# access-control vmap 21 mirror port 4
CN 4093(config)# access-control vmap 21 action permit
CN 4093(config)# vlan 3
CN 4093(config-vlan)# vmap 21 intports
```

Management ACLs

Management ACLs (MACLs) filter inbound traffic (traffic heading toward the CPU). MACLs are applied switch-wide. Traffic can be filtered based on the following:

- IPv4 source address
- IPv4 destination address
- IPv4 protocols
- TCP/UDP destination or source port

Lower MACL numbers have higher priority. Up to 128 MACLs can be configured.

Following is an example MACL configuration based on a destination IP address and a TCP-UDP destination port:

```
CN 4093(config)# access-control macl 1 ipv4 destination-ip-address
1.1.1.1 255.255.255.0
CN 4093(config)# access-control macl 1 tcp-udp destination-port 111
0xffff
CN 4093(config)# access-control macl 1 statistics
CN 4093(config)# access-control macl 1 action permit
CN 4093(config)# access-control macl 1 enable
```

Use the following command to view the MACL configuration:

```
CN 4093(config)# show access-control macl 1

MACL 1 profile  : Enabled
   IPv4
     - DST IP     : 1.1.1.1/255.255.255.0
   TCP/UDP
     - DST Port   : 111/0xffff
   Action       : Permit
   Statistics   : Enabled
```

Part 3: Switch Basics

This section discusses basic switching functions:

- VLANs
- Port Aggregation
- Spanning Tree Protocols (Spanning Tree Groups, Rapid Spanning Tree Protocol and Multiple Spanning Tree Protocol)
- Quality of Service

Chapter 8. VLANs

This chapter describes network design and topology considerations for using Virtual Local Area Networks (VLANs). VLANs are commonly used to split up groups of network users into manageable broadcast domains, to create logical segmentation of workgroups, and to enforce security policies among logical segments. The following topics are discussed in this chapter:

Note: Basic VLANs can be configured during initial switch configuration (see "Using the Setup Utility" in the *CN4093 Enterprise NOS 8.4 Command Reference*). More comprehensive VLAN configuration can be done from the Command Line Interface (see "VLAN Configuration" as well as "Port Configuration" in the *CN4093 Enterprise NOS 8.4 Command Reference*).

VLANs Overview

Setting up virtual LANs (VLANs) is a way to segment networks to increase network flexibility without changing the physical network topology. With network segmentation, each switch port connects to a segment that is a single broadcast domain. When a switch port is configured to be a member of a VLAN, it is added to a group of ports (workgroup) that belong to one broadcast domain.

Ports are grouped into broadcast domains by assigning them to the same VLAN. Frames received in one VLAN can only be forwarded within that VLAN, and multicast, broadcast, and unknown unicast frames are flooded only to ports in the same VLAN.

The CN4093 automatically supports jumbo frames. This default cannot be manually configured or disabled.

The CN4093 10 Gb Converged Scalable Switch (CN4093) supports jumbo frames with a Maximum Transmission Unit (MTU) of 9,216 bytes. Within each frame, 18 bytes are reserved for the Ethernet header and CRC trailer. The remaining space in the frame (up to 9,198 bytes) comprise the packet, which includes the payload of up to 9,000 bytes and any additional overhead, such as 802.1q or VLAN tags. Jumbo frame support is automatic: it is enabled by default, requires no manual configuration, and cannot be manually disabled.

Note: Jumbo frames are not supported for traffic sent to switch management interfaces.

VLANs and Port VLAN ID Numbers

VLAN Numbers

Enterprise NOS supports up to 4095 VLANs per switch. Even though the maximum number of VLANs supported at any given time is 4095, each can be identified with any number between 1 and 4094. VLAN 1 is the default VLAN for the external ports and the internal blade ports.

VLAN 4095 is reserved for use by the management network, which includes the management ports and (by default) internal ports. This configuration allows Serial over LAN (SoL) management—a feature available on certain server blades. Management functions can also be assigned to other VLANs (using the following command:

```
CN 4093(config)# vlan <1-4094>
CN 4093(config-vlan)# management
```

Use the following command to view VLAN information:

```
CN 4093# show vlan

VLAN           Name                Status MGT           Ports
----  --------------------------  ------ --- ------------------------
1     Default VLAN                ena    dis  INTA1-EXT22
200   VLAN 200                    dis    dis  empty
300   VLAN 300                    dis    dis  empty
4095  Mgmt VLAN                   ena    ena  EXTM MGT1

Primary  Secondary  Type             Ports                 vPorts
-------  ---------  ---------------  --------------------  ---------
```

Note: The sample screens that appear in this document might differ slightly from the screens displayed by your system. Screen content varies based on the type of blade chassis unit that you are using and the firmware versions and options that are installed.

PVID/Native VLAN Numbers

Each port in the switch has a configurable default VLAN number, known as its *PVID*. By default, the PVID for all non-management ports is set to 1, which correlates to the default VLAN ID. The PVID for each port can be configured to any VLAN number between 1 and 4094.

Use the following CLI commands to view PVIDs:

- Port information:

```
CN 4093# show interface information
(or)
CN 4093# show interface trunk

Alias   Port Tag    Type    RMON Lrn Fld PVID     DESCRIPTION  VLAN(s)
             Trk                         NVLAN
------- ---- --- ---------- ---- --- --- ------ -------------- ------
INTA1   1     n  Internal    d    e   e    4094   INTA1          4094
INTA2   2     n  Internal    d    e   e       1   INTA2          1
INTA3   3     n  Internal    d    e   e       1   INTA3          1
INTA4   4     n  Internal    d    e   e       1   INTA4          1
INTA5   5     n  Internal    d    e   e       1   INTA5          1
INTA6   6     n  Internal    d    e   e       1   INTA6          1
INTA7   7     n  Internal    d    e   e       1   INTA7          1
INTA8   8     n  Internal    d    e   e       1   INTA8          1
INTA9   9     n  Internal    d    e   e       1   INTA9          1
...
INTC13  41    n  Internal    d    e   e       1   INTC13         1
INTC14  42    n  Internal    d    e   e       1   INTC14         1
EXT1    43    n  External    d    e   e       1   EXT1           1
EXT2    44    n  External    d    d   d       1   EXT2           1
EXT3    45    n  External    d    d   d       1   EXT3           1
EXT4    46    n  External    d    e   d       1   EXT4           1
EXT5    47    n  External    d    d   d       1   EXT5           1
EXT6    48    n  External    d    e   e       1   EXT6           1
EXT7    49    n  External    d    e   e       1   EXT7           1
EXT8    50    n  External    d    e   e       1   EXT8           1
EXT9    51    n  External    d    e   e       1   EXT9           1
EXT10   52    n  External    d    e   e       1   EXT10          1
EXT11   53    n  External    d    e   e       1   EXT11          1
EXT12   54    n  External    d    e   e       1   EXT12          1
EXT13   55    n  External    d    e   e       1   EXT13          1
EXT14   56    n  External    d    e   e       1   EXT14          1
EXT15   57    n  External    d    d   d       1   EXT15          1
EXT19   61    n  External    d    e   e       1   EXT19          1
EXTM    65    n  Mgmt        d    e   e    4095   EXTM           4095
MGT1    66    y  Mgmt        d    e   e    4095   MGT1           4095

* = PVID/Native-VLAN is tagged.
# = PVID is ingress tagged.
Trk = Trunk mode
NVLAN = Native-VLAN
```

Note: The sample output that appears in this document might differ slightly from that displayed by your system. Output varies based on the type of blade chassis unit that you are using and the firmware versions and options that are installed.

- Port Configuration:
  - Access mode port:

```
CN 4093(config)# interface port <port number>
CN 4093(config-if)# switchport access vlan <VLAN ID>
```

  - Trunk mode port:

```
CN 4093(config)# interface port <port number>
CN 4093(config-if)# switchport trunk native vlan <VLAN ID>
```

Each port on the switch can belong to one or more VLANs, and each VLAN can have any number of switch ports in its membership. Any port that belongs to multiple VLANs, however, must have VLAN *tagging* enabled (see "VLAN Tagging/Trunk Mode" on page 146).

VLAN Tagging/Trunk Mode

Enterprise NOS software supports 802.1Q VLAN *tagging*, providing standards-based VLAN support for Ethernet systems.

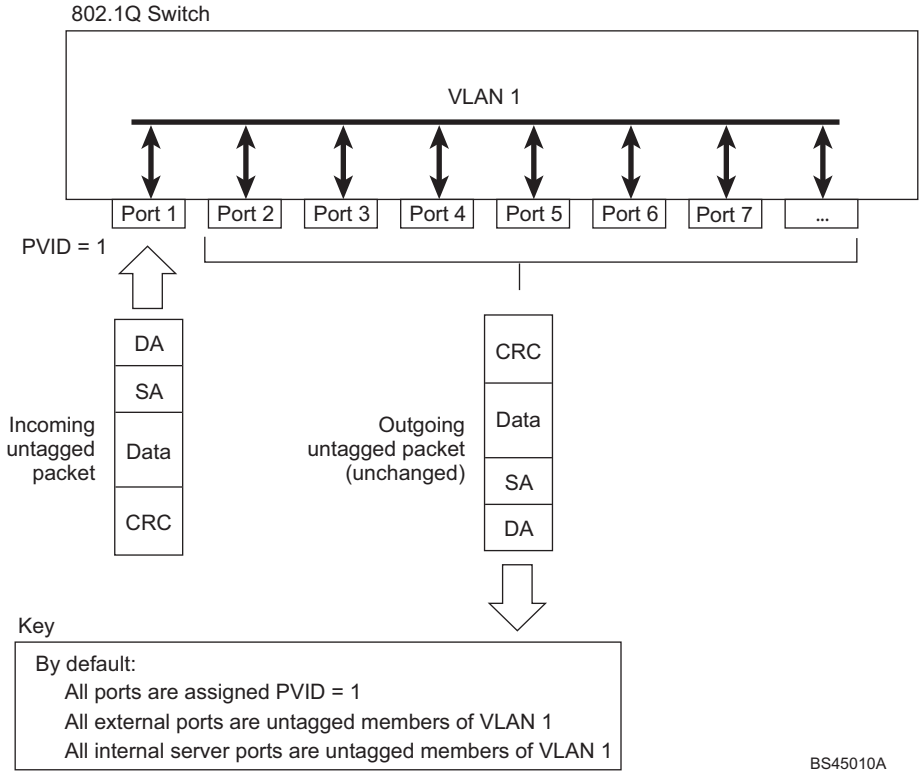
Tagging places the VLAN identifier in the frame header of a packet, allowing each port to belong to multiple VLANs. When you add a port to multiple VLANs, you also must enable tagging on that port.

Since tagging fundamentally changes the format of frames transmitted on a tagged port, you must carefully plan network designs to prevent tagged frames from being transmitted to devices that do not support 802.1Q VLAN tags, or devices where tagging is not enabled.

Important terms used with the 802.1Q tagging feature are:

- VLAN identifier (VID)—the 12-bit portion of the VLAN tag in the frame header that identifies an explicit VLAN.
- Port VLAN identifier (PVID)—a classification mechanism that associates a port with a specific VLAN. For example, a port with a PVID of 3 (PVID =3) assigns all untagged frames received on this port to VLAN 3. Any untagged frames received by the switch are classified with the PVID of the receiving port.
- Tagged frame—a frame that carries VLAN tagging information in the header. This VLAN tagging information is a 32-bit field (VLAN tag) in the frame header that identifies the frame as belonging to a specific VLAN. Untagged frames are marked (tagged) with this classification as they leave the switch through a port that is configured as a tagged port.
- Untagged frame— a frame that does not carry any VLAN tagging information in the frame header.
- Untagged member—a port that has been configured as an untagged member of a specific VLAN. When an untagged frame exits the switch through an untagged member port, the frame header remains unchanged. When a tagged frame exits the switch through an untagged member port, the tag is stripped and the tagged frame is changed to an untagged frame.
- Tagged member—a port that has been configured as a tagged member of a specific VLAN. When an untagged frame exits the switch through a tagged member port, the frame header is modified to include the 32-bit tag associated with the PVID. When a tagged frame exits the switch through a tagged member port, the frame header remains unchanged (original VID remains).

Figure 2. Default VLAN settings



Note: The port numbers specified in these illustrations may not directly correspond to the physical port configuration of your switch model.

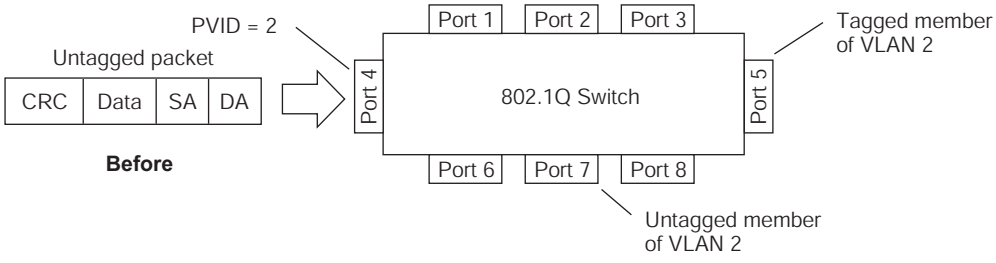
When a VLAN is configured, ports are added as members of the VLAN, and the ports are defined as either *tagged* or *untagged* (see Figure 3 through Figure 6).

The default configuration settings for CN4093s have all ports set as untagged members of VLAN 1 with all ports configured as PVID = 1. In the default configuration example shown in Figure 2, all incoming packets are assigned to VLAN 1 by the default port VLAN identifier (PVID =1).

Figure 3 through Figure 6 illustrate generic examples of VLAN tagging. In Figure 3, untagged incoming packets are assigned directly to VLAN 2 (PVID = 2). Port 5 is configured as a *tagged* member of VLAN 2, and port 7 is configured as an *untagged* member of VLAN 2.

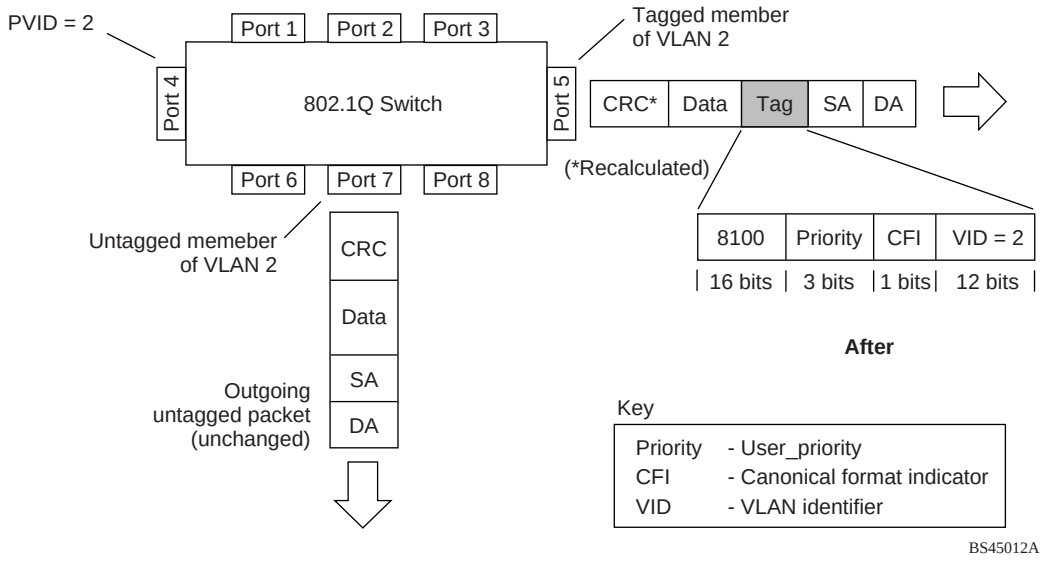
Note: The port assignments in the following figures are general examples and are not meant to match any specific CN4093.

Figure 3. Port-based VLAN assignment



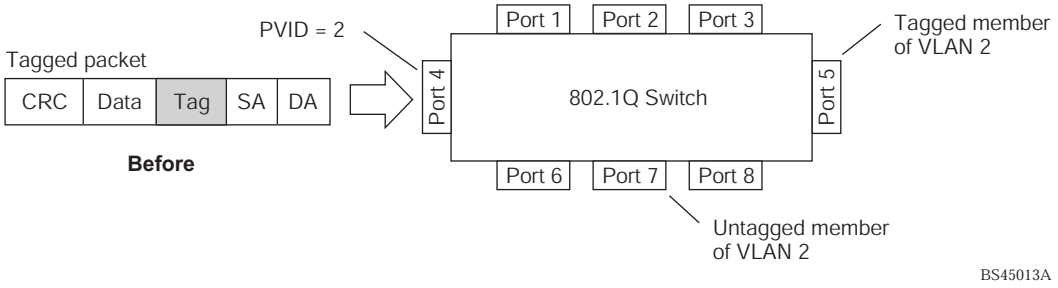
As shown in Figure 4, the untagged packet is marked (tagged) as it leaves the switch through port 5, which is configured as a tagged member of VLAN 2. The untagged packet remains unchanged as it leaves the switch through port 7, which is configured as an untagged member of VLAN 2.

Figure 4. 802.1Q tagging (after port-based VLAN assignment)



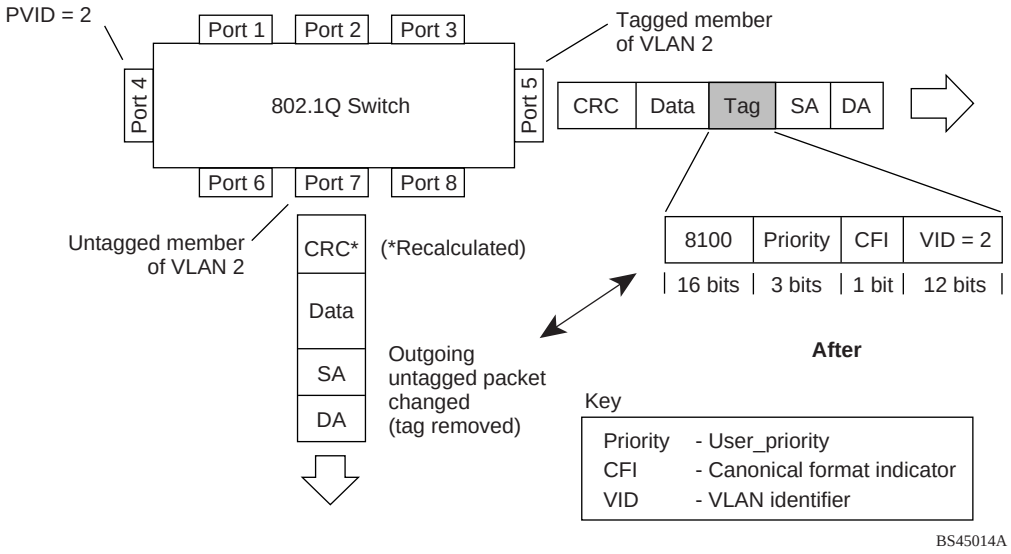
In Figure 5, tagged incoming packets are assigned directly to VLAN 2 because of the tag assignment in the packet. Port 5 is configured as a *tagged* member of VLAN 2 and port 7 is configured as an *untagged* member of VLAN 2.

Figure 5. 802.1Q tag assignment



As shown in Figure 6, the tagged packet remains unchanged as it leaves the switch through port 5, which is configured as a tagged member of VLAN 2. However, the tagged packet is stripped (untagged) as it leaves the switch through port 7, which is configured as an untagged member of VLAN 2.

Figure 6. 802.1Q tagging (after 802.1Q tag assignment)



Note: Setting the configuration to factory default (`CN 4093(config)#` **`boot configuration-block factory`**) will reset all non-management ports to VLAN 1.

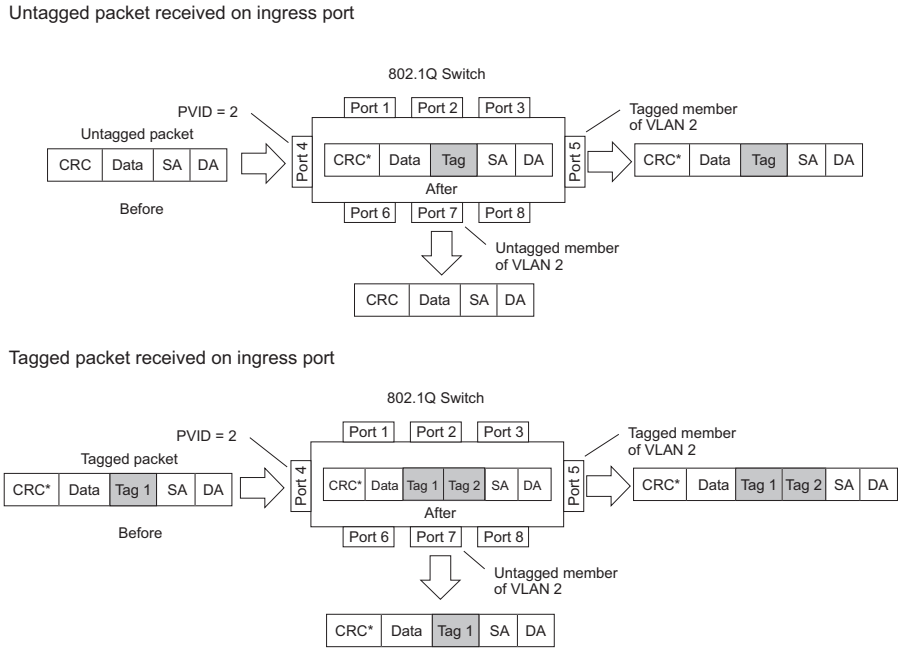
Ingress VLAN Tagging

Tagging can be enabled on an ingress port. When a packet is received on an ingress port, and if ingress tagging is enabled on the port, a VLAN tag with the port PVID is inserted into the packet as the outer VLAN tag. Depending on the egress port setting (tagged or untagged), the outer tag of the packet is retained or removed when it leaves the egress port.

Ingress VLAN tagging is used to tunnel packets through a public domain without altering the original 802.1Q status.

When ingress tagging is enabled on a port, all packets, whether untagged or tagged, will be tagged again. As shown in Figure 7, when tagging is enabled on the egress port, the outer tag of the packet is retained when it leaves the egress port. If tagging is disabled on the egress port, the outer tag of the packet is removed when it leaves the egress port.

Figure 7. 802.1Q tagging (after ingress tagging assignment)



By default, ingress tagging is disabled. To enable ingress tagging on a port, use the following commands:

```
CN 4093(config)# interface port <number>
CN 4093(config-if)# tagpvid-ingress
```

Limitations

Ingress tagging cannot be configured with the following features/configurations:

- vNIC ports
- VMready ports
- UFP ports
- Management ports

VLAN Topologies and Design Considerations

- By default, the Enterprise NOS software is configured so that tagging is disabled on all external ports and on all internal ports.
- By default, the Enterprise NOS software is configured so that all internal ports are members of VLAN 1.
- By default, the Enterprise NOS software is configured so that the management port is a member of the default management VLAN 4095.
- Multiple management VLANs can be configured on the switch, in addition to the default VLAN 4095, using the following commands:

```
CN 4093(config)# vlan <VLAN ID (1-4094)>
CN 4093(config-vlan)# management
```

- When using Spanning Tree, STG 2-128 may contain only one VLAN unless Multiple Spanning-Tree Protocol (MSTP) mode is used. With MSTP mode, STG 1 to 32 can include multiple VLANs.

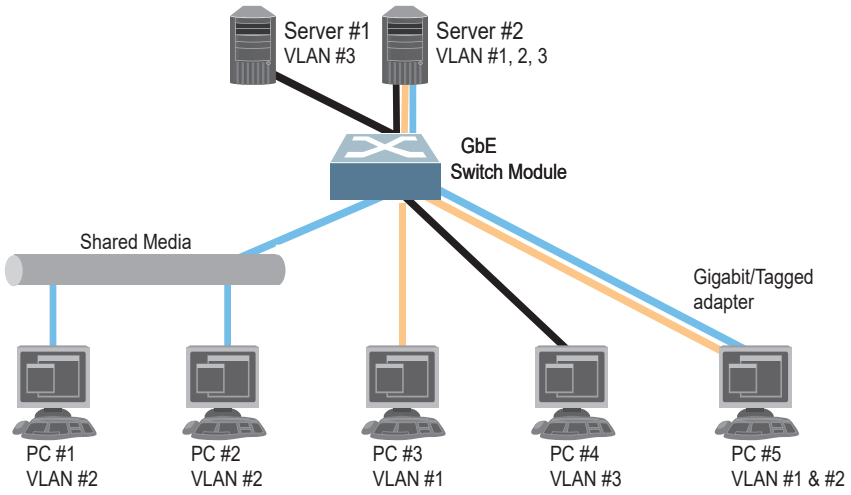
VLAN Configuration Rules

VLANs operate according to specific configuration rules. When creating VLANs, consider the following rules that determine how the configured VLAN reacts in any network topology:

- All ports involved in aggregation and port mirroring must have the same VLAN configuration. If a port is on a LAG with a mirroring port, the VLAN configuration cannot be changed. For more information about aggregation, see "Configuring a Static LAG" on page 167.
- If a port is configured for port mirroring, the port's VLAN membership cannot be changed. For more information on configuring port mirroring, see "Port Mirroring" on page 611.
- Management VLANs must contain the management port, and can include one or more internal ports (INT*x*). External ports (EXT*x*) cannot be members of any management VLAN.

Example: Multiple VLANs with Tagging Adapters

Figure 8. Multiple VLANs with VLAN-Tagged Gigabit Adapters



The features of this VLAN are described in the following table:

Component	Description
Switch	This switch is configured for three VLANs that represent three different IP subnets. Two servers and five clients are attached to the switch.
Server #1	This server is a member of VLAN 3 and has presence in only one IP subnet. The associated internal switch port is only a member of VLAN 3, so tagging is disabled.
Server #2	This high-use server needs to be accessed from all VLANs and IP subnets. The server has a VLAN-tagging adapter installed with VLAN tagging turned on. The adapter is attached to one of the internal switch ports, that is a member of VLANs 1, 2, and 3, and has tagging enabled. Because of the VLAN tagging capabilities of both the adapter and the switch, the server is able to communicate on all three IP subnets in this network. Broadcast separation between all three VLANs and subnets, however, is maintained.
PCs #1 and #2	These PCs are attached to a shared media hub that is then connected to the switch. They belong to VLAN 2 and are logically in the same IP subnet as Server 2 and PC 5. The associated external switch port has tagging disabled.
PC #3	A member of VLAN 1, this PC can only communicate with Server 2 and PC 5. The associated external switch port has tagging disabled.

Component	Description
PC #4	A member of VLAN 3, this PC can only communicate with Server 1 and Server 2. The associated external switch port has tagging disabled.
PC #5	A member of both VLAN 1 and VLAN 2, this PC has a VLAN-tagging Gigabit Ethernet adapter installed. It can communicate with Server 2 and PC 3 via VLAN 1, and to Server 2, PC 1 and PC 2 via VLAN 2. The associated external switch port is a member of VLAN 1 and VLAN 2, and has tagging enabled.

Note: VLAN tagging is required only on ports that are connected to other CN4093s or on ports that connect to tag-capable end-stations, such as servers with VLAN-tagging adapters.

Protocol-Based VLANs

Protocol-based VLANs (PVLANs) allow you to segment network traffic according to the network protocols in use. Traffic for supported network protocols can be confined to a particular port-based VLAN. You can give different priority levels to traffic generated by different network protocols.

With PVLAN, the switch classifies incoming packets by Ethernet protocol of the packets, not by the configuration of the ingress port. When an untagged or priority-tagged frame arrives at an ingress port, the protocol information carried in the frame is used to determine a VLAN to which the frame belongs. If a frame's protocol is not recognized as a pre-defined PVLAN type, the ingress port's PVID is assigned to the frame. When a tagged frame arrives, the VLAN ID in the frame's tag is used.

Each VLAN can contain up to eight different PVLANs. You can configure separate PVLANs on different VLANs, with each PVLAN segmenting traffic for the same protocol type. For example, you can configure PVLAN 1 on VLAN 2 to segment IPv4 traffic, and PVLAN 8 on VLAN 100 to segment IPv4 traffic.

To define a PVLAN on a VLAN, configure a PVLAN number (1-8) and specify the frame type and the Ethernet type of the PVLAN protocol. You must assign at least one port to the PVLAN before it can function. Define the PVLAN frame type and Ethernet type as follows:

- Frame type—consists of one of the following values:
  - Ether2 (Ethernet II)
  - SNAP (Subnetwork Access Protocol)
  - LLC (Logical Link Control)
- Ethernet type—consists of a 4-digit (16 bit) hex value that defines the Ethernet type. You can use common Ethernet protocol values, or define your own values. Following are examples of common Ethernet protocol values:
  - IPv4 = `0800`
  - IPv6 = `86dd`
  - ARP = `0806`

Port-Based vs. Protocol-Based VLANs

Each VLAN supports both port-based and protocol-based association, as follows:

- The default VLAN configuration is port-based. All data ports are members of VLAN 1, with no PVLAN association.
- When you add ports to a PVLAN, the ports become members of both the port-based VLAN and the PVLAN. For example, if you add port EXT1 to PVLAN 1 on VLAN 2, the port also becomes a member of VLAN 2.
- When you delete a PVLAN, it's member ports remain members of the port-based VLAN. For example, if you delete PVLAN 1 from VLAN 2, port EXT1 remains a member of VLAN 2.
- When you delete a port from a VLAN, the port is deleted from all corresponding PVLANs.

PVLAN Priority Levels

You can assign each PVLAN a priority value of 0-7, used for Quality of Service (QoS). PVLAN priority takes precedence over a port's configured priority level. If no priority level is configured for the PVLAN (priority = 0), each port's priority is used (if configured).

All member ports of a PVLAN have the same PVLAN priority level.

PVLAN Tagging

When PVLAN tagging is enabled, the switch tags frames that match the PVLAN protocol. For more information about tagging, see "VLAN Tagging/Trunk Mode" on page 146.

Untagged ports must have PVLAN tagging disabled. Tagged ports can have PVLAN tagging either enabled or disabled.

PVLAN tagging has higher precedence than port-based tagging. If a port is tag enabled, and the port is a member of a PVLAN, the PVLAN tags egress frames that match the PVLAN protocol.

Use the tag-pvlan command (**`vlan`** <*x*> **`protocol-vlan`** <*x*> **`tag-pvlan`** <*x*>) to define the complete list of tag-enabled ports in the PVLAN. Note that all ports not included in the PVLAN tag list will have PVLAN tagging disabled.

PVLAN Configuration Guidelines

Consider the following guidelines when you configure protocol-based VLANs:

- Each port can support up to 8 VLAN protocols.
- The CN4093 can support up to 16 protocols simultaneously.
- Each PVLAN must have at least one port assigned before it can be activated.
- The same port within a port-based VLAN can belong to multiple PVLANs.
- An untagged port can be a member of multiple PVLANs.
- A port cannot be a member of different VLANs with the same protocol association.

Configuring PVLAN

Follow this procedure to configure a Protocol-based VLAN (PVLAN).

1. Configure VLAN tagging/trunk mode for ports.

```
CN 4093(config)# interface port 1,2
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit
```

2. Create a VLAN and define the protocol type(s) supported by the VLAN.

```
CN 4093(config)# vlan 2
CN 4093(config-vlan)# protocol-vlan 1 frame-type ether2 0800
```

3. Configure the priority value for the protocol.

```
CN 4093(config-vlan)# protocol-vlan 1 priority 2
```

4. Add member ports for this PVLAN.

```
CN 4093(config-vlan)# protocol-vlan 1 member 1,2
```

Note: If VLAN tagging is turned on and the port being added to the VLAN has a different default VLAN (PVID/Native VLAN), you will be asked to confirm changing the PVID to the current VLAN.

5. Enable the PVLAN.

```
CN 4093(config-vlan)# protocol-vlan 1 enable
CN 4093(config-vlan)# exit
```

6. Verify PVLAN operation.

```
CN 4093(config)# show vlan

VLAN                Name            Status MGT           Ports
----  ----------------------------  ------ --- -------------------------
1     Default VLAN                  ena    dis  INTA2-EXT15 EXT19
2     VLAN 2                        ena    dis  INTA1 INTA2

Primary  Secondary  Type         Ports               vPorts
-------  ---------  -------  ----------------  ------------------

PVLAN  Protocol  FrameType  EtherType  Priority  Status     Ports
-----  --------  ---------  ---------  --------  -------   --------------
2      1         Ether2     0800       2         enabled   INTA1 INTA2

PVLAN      PVLAN-Tagged Ports
-----   --------------------------
none    none
```

Private VLANs

Private VLANs provide Layer 2 isolation between the ports within the same broadcast domain. Private VLANs can control traffic within a VLAN domain, and provide port-based security for host servers.

Enterprise NOS supports Private VLAN configuration as described in RFC 5517.

Use Private VLANs to partition a VLAN domain into sub-domains. Each sub-domain is comprised of one primary VLAN and one secondary VLAN, as follows:

- Primary VLAN—carries unidirectional traffic downstream from promiscuous ports. Each Private VLAN has only one primary VLAN. All ports in the Private VLAN are members of the primary VLAN.
- Secondary VLAN—Secondary VLANs are internal to a private VLAN domain, and are defined as follows:
  - Isolated VLAN—carries unidirectional traffic upstream from the host servers toward ports in the primary VLAN. Each Private VLAN can contain only one Isolated VLAN.
  - Community VLAN—carries upstream traffic from ports in the community VLAN to other ports in the same community, and to ports in the primary VLAN. Each Private VLAN can contain multiple community VLANs.

After you define the primary VLAN and one or more secondary VLANs, you map the secondary VLAN(s) to the primary VLAN.

Private VLAN Ports

Private VLAN ports are defined as follows:

- Promiscuous—A promiscuous port is a port that belongs to the primary VLAN. The promiscuous port can communicate with all the interfaces, including ports in the secondary VLANs (Isolated VLAN and Community VLANs).
- Isolated—An isolated port is a host port that belongs to an isolated VLAN. Each isolated port has complete layer 2 separation from other ports within the same private VLAN (including other isolated ports), except for the promiscuous ports.
  - Traffic sent to an isolated port is blocked by the Private VLAN, except the traffic from promiscuous ports.
  - Traffic received from an isolated port is forwarded only to promiscuous ports.
- Community—A community port is a host port that belongs to a community VLAN. Community ports can communicate with other ports in the same community VLAN, and with promiscuous ports. These interfaces are isolated at layer 2 from all other interfaces in other communities and from isolated ports within the Private VLAN.

Configuration Guidelines

The following guidelines apply when configuring Private VLANs:

- Management VLANs cannot be Private VLANs. Management ports cannot be members of a Private VLAN.
- The default VLAN 1 cannot be a Private VLAN.
- IGMP Snooping must be disabled on Private VLANs.
- All VLANs that comprise the Private VLAN must belong to the same Spanning Tree Group.
- A VLAN pair is a primary VLAN and one associated secondary VLAN (isolated or community). The maximum number of VLAN pairs per port is 16.

Configuration Example

Follow this procedure to configure a Private VLAN.

1. Select a VLAN and define the Private VLAN type as primary.

```
CN 4093(config)# vlan 700
CN 4093(config-vlan)# private-vlan primary
CN 4093(config-vlan)# exit
```

2. Configure a promiscuous port for VLAN 700.

```
CN 4093(config)# interface port 1
CN 4093(config-if)# switchport mode private-vlan
CN 4093(config-if)# switchport private-vlan mapping 700
CN 4093(config-if)# exit
```

3. Configure two secondary VLANs: isolated VLAN and community VLAN.

```
CN 4093(config)# vlan 701
CN 4093(config-vlan)# private-vlan isolated
CN 4093(config-vlan)# exit
CN 4093(config)# vlan 702
CN 4093(config-vlan)# private-vlan community
CN 4093(config-vlan)# exit
```

4. Map secondary VLANs to primary VLAN.

```
CN 4093(config)# vlan 700-702
CN 4093(config-vlan)# stg 1
CN 4093(config-vlan)# exit
CN 4093(config)# vlan 700
CN 4093(config-vlan)# private-vlan association 701,702
CN 4093(config-vlan)# exit
```

5. Configure host ports for secondary VLANs.

```
CN 4093(config)# interface port 2
CN 4093(config-if)# switchport mode private-vlan
CN 4093(config-if)# switchport private-vlan host-association 700 701
CN 4093(config-if)# exit

CN 4093(config)# interface port 3
CN 4093(config-if)# switchport mode private-vlan
CN 4093(config-if)# switchport private-vlan host-association 700 702
CN 4093(config-if)# exit
```

6. Verify the configuration.

```
CN 4093(config)# show vlan private-vlan

Primary  Secondary      Type                Ports
-------  --------- --------------- --------------------------------
700         701        isolated             1 2
700         702        community            1 3
```

Chapter 9. Ports and Link Aggregation (LAG)

LAGs can provide super-bandwidth, multi-link connections between the CN4093 10 Gb Converged Scalable Switch (CN4093) and other LAG-capable devices. A LAG is a group of ports that act together, combining their bandwidth to create a single, larger virtual link. This chapter provides configuration background and examples for aggregating multiple ports together:

Configuring Port Modes

The switch allows you to set the port mode. Select the port mode that fits your network configuration.

Switch port modes are available based on the installation license.

The following port modes are available:

- **Base Port** mode:
  - Fourteen 10Gb internal (1 port x 14 blade servers)
  - Eight 10Gb external
- **Upgrade 1 Port** mode:
  - Twenty Eight 10Gb internal (2 ports x 14 blade servers)
  - Eight 10Gb external
  - Two 40Gb external
- **Upgrade 2 Port** mode:
  - Forty Two 10Gb internal (3 ports x 14 Blade servers)
  - Fourteen 10Gb external
  - Two 40Gb external

Base Port mode is the default. To upgrade the port mode, you must obtain a software license key.

The following command sequence is an example of how to upgrade the port mode (e.g. switch SN Y010CM2CN058):

```
CN 4093# software-key
Enter hostname or IP address of SFTP/TFTP server: 9.44.143.105
Enter name of file on SFTP/TFTP server:
ibm_fod_0019_Y010CM2CN058_anyos_noarch.key
Enter username for SFTP server or hit return for TFTP server:
Enter the port to use for downloading the file
["data"|"extm"|"mgt"]:  mgt

Starting download key file...
Key file download complete (502 bytes)
Software feature 'Upgrade1' will be Active upon next reboot.
NOTICE  mgmt: Software feature 'Upgrade1' will be Active upon next
reboot.

A Reboot is required for the new settings to take effect.
```

Note: Upgrade 1 and Upgrade 2 can be independently installed in any order. You can choose to install any one upgrade or both.

After you enter the license key, you must reset the switch (`CN 4093# reload`) for the change to take affect. Use the following command to verify the port configuration:

```
CN 4093(config)# show interface information

Alias Port Tag    Type    RMON Lrn Fld PVID     DESCRIPTION     VLAN(s)
           Trk                       NVLAN
----- ---- --- ---------- ---- --- --- ------  ------------  -----------
INTA1   1    n  Internal    d    e   e     1   INTA1          1
INTA2   2    n  Internal    d    e   e     1   INTA2          1
INTA3   3    n  Internal    d    e   e     1   INTA3          1
INTA4   4    n  Internal    d    e   e     1   INTA4          1
INTA5   5    n  Internal    d    e   e     1   INTA5          1
INTA6   6    n  Internal    d    e   e     1   INTA6          1
INTA7   7    n  Internal    d    e   e     1   INTA7          1
INTA8   8    n  Internal    d    e   e     1   INTA8          1
INTA9   9    n  Internal    d    e   e     1   INTA9          1
...
EXT1    43   n  External    d    e   e     1   EXT1           1
EXT2    44   n  External    d    e   e     1   EXT2           1
EXT3    45   n  External    d    e   e     1   EXT3           1
EXT4    46   n  External    d    e   e     1   EXT4           1
EXT5    47   n  External    d    e   e     1   EXT5           1
EXT6    48   n  External    d    e   e     1   EXT6           1
EXT7    49   n  External    d    e   e     1   EXT7           1
EXT8    50   n  External    d    e   e     1   EXT8           1
EXT9    51   n  External    d    e   e     1   EXT9           1
EXT10   52   n  External    d    e   e     1   EXT10          1
EXT11   53   n  External    d    e   e     1   EXT11          1
EXT12   54   n  External    d    e   e     1   EXT12          1
EXT13   55   n  External    d    e   e     1   EXT13          1
EXT14   56   n  External    d    e   e     1   EXT14          1
EXT15   57   n  External    d    e   e     1   EXT15          1
EXT16   58   n  External    d    e   e     1   EXT16          1
EXT17   59   n  External    d    e   e     1   EXT17          1
EXT18   60   n  External    d    e   e     1   EXT18          1
EXT19   61   n  External    d    e   e     1   EXT19          1
EXT20   62   n  External    d    e   e     1   EXT20          1
EXT21   63   n  External    d    e   e     1   EXT21          1
EXT22   64   n  External    d    e   e     1   EXT22          1
EXTM    65   n  Mgmt        d    e   e  4095   EXTM           4095
MGT1    66   y  Mgmt        d    e   e  4095   MGT1           4095

* = PVID/Native-VLAN is tagged.
# = PVID is ingress tagged.
Trk  = Trunk mode
NVLAN = Native-VLAN
```

Configuring QSFP+ Ports

QSFP+ ports support both 10GbE and 40GbE, as shown in Table 15.

Table 15. *QSFP+ Port Numbering*

Physical Port Number	40GbE mode	10GbE mode
Port EXT3	Port EXT3	Ports EXT3-EXT6
Port EXT7	Port EXT7	Ports EXT7-EXT10

QSFP+ ports are available only when Upgrade 1 is installed (see "Configuring Port Modes" on page 162).

The following procedure allows you to change the QSFP+ port mode.

1. Display the current port mode for the QSFP+ ports.

```
CN 4093# show boot qsfp-port-modes

QSFP ports booted configuration:
  Port EXT3, EXT4, EXT5, EXT6 - 10G Mode
  Port EXT7, EXT8, EXT9, EXT10 - 10G Mode

QSFP ports saved configuration:
  Port EXT3, EXT4, EXT5, EXT6 - 10G Mode
  Port EXT7, EXT8, EXT9, EXT10 - 10G Mode
```

2. Change the port mode to 40GbE. Select the physical port number.

```
CN 4093(config)# boot qsfp-40Gports ext3
```

3. Verify the change.

```
CN 4093# show boot qsfp-port-modes

QSFP ports booted configuration:
  Port EXT3, EXT4, EXT5, EXT6 - 10G Mode
  Port EXT7, EXT8, EXT9, EXT10 - 10G Mode

QSFP ports saved configuration:
  Port EXT3 - 40G Mode
  Port EXT7, EXT8, EXT9, EXT10 - 10G Mode
```

4. Reset the switch.

```
CN 4093(config)# reload
```

Remove the configured port from QSFP+ mode to reset the ports to 10GbE mode.

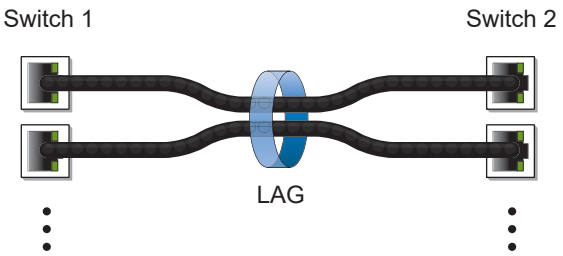
```
CN 4093(config)# no boot qsfp-40Gports ext3
```

Aggregation Overview

When using LAGs between two switches, as shown in Figure 9, you can create a virtual link between them, operating with combined throughput levels that depends on how many physical ports are included.

Two types of aggregation are available: static LAGs and dynamic Link Aggregation Control Protocol (LACP) LAGs. Up to 52 LAGs of each type are supported, depending of the number and type of available ports. Each LAG can include up to 24 member ports.

Figure 9. Link Aggregation Group (LAG)



LAGs are also useful for connecting a CN4093 to third-party devices that support link aggregation, such as Cisco routers and switches with EtherChannel technology (*not* ISL aggregation technology) and Sun's Quad Fast Ethernet Adapter. Static LAG technology is compatible with these devices when they are configured manually.

LAG traffic is statistically distributed among the ports in a LAG, based on a variety of configurable options.

Also, since each LAG is comprised of multiple physical links, the LAG is inherently fault tolerant. As long as one connection between the switches is available, the LAG remains active and statistical load balancing is maintained whenever a port in the LAG is lost or returned to service.

Static LAGs

When you create and enable a static LAG, the LAG members (switch ports) take on certain settings necessary for correct operation of the aggregation feature.

Before Configuring Static LAGs

Before you configure your LAG, you must consider these settings, along with specific configuration rules, as follows:

- Read the configuration rules provided in the section, "Static LAG Configuration Rules" on page 166."
- Determine which switch ports are to become *LAG members* (the specific ports making up the LAG).
- Ensure that the chosen switch ports are set to `enabled`.
- Ensure all member ports in a LAG have the same VLAN configuration.
- Consider how the existing Spanning Tree will react to the new LAG configuration. See "Spanning Tree Protocols" on page 175 for configuration guidelines.
- Consider how existing VLANs will be affected by the addition of a LAG.

Static LAG Configuration Rules

The aggregation feature operates according to specific configuration rules. When creating LAGs, consider the following rules that determine how a LAG reacts in any network topology:

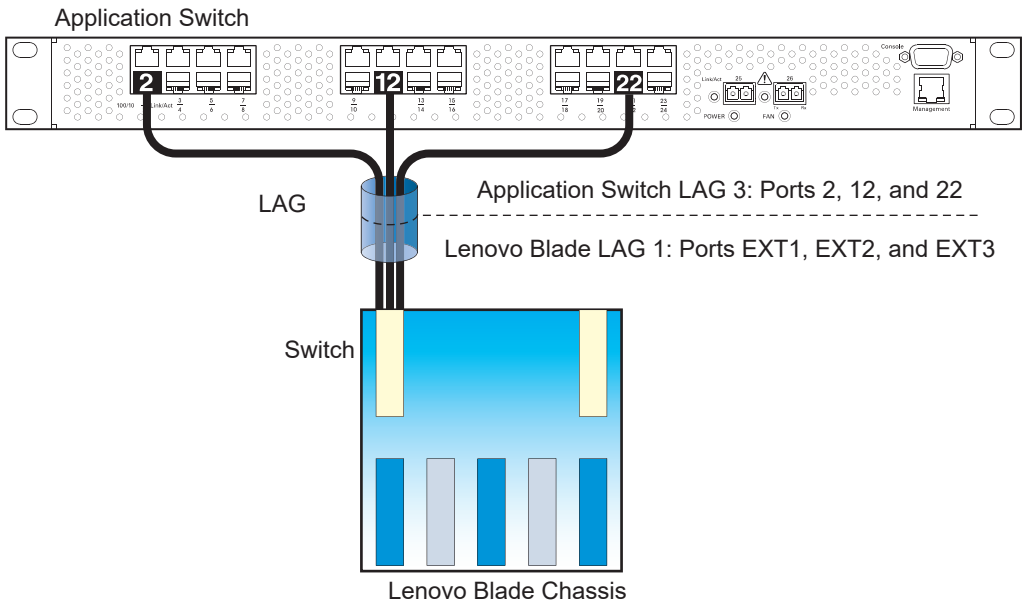
- All LAGs must originate from one network entity (a single device or multiple devices acting in a stack) and lead to one destination entity. For example, you cannot combine links from two different servers into one LAG.
- Any physical switch port can belong to only one LAG.
- Depending on port availability, the switch supports up to 24 ports in each LAG.
- Internal ports (INT*x*) and external ports (EXT*x*) cannot become members of the same LAG.
- Aggregation from third-party devices must comply with Cisco® EtherChannel® technology.
- All LAG member ports must be assigned to the same VLAN configuration before the LAG can be enabled.
- If you change the VLAN settings of any LAG member, you cannot apply the change until you change the VLAN settings of all LAG members.
- When an active port is configured in a LAG, the port becomes a *LAG member* when you enable the LAG. The Spanning Tree parameters for the port then change to reflect the new LAG settings.
- All LAG members must be in the same Spanning Tree Group (STG) and can belong to only one Spanning Tree Group (STG). However if all ports are *tagged*, then all LAG ports can belong to multiple STGs.

- If you change the Spanning Tree participation of any LAG member to `enabled` or `disabled`, the Spanning Tree participation of all members of that LAG should be changed similarly.
- When a LAG is enabled, the LAG's Spanning Tree participation setting takes precedence over that of any LAG member.
- You cannot configure a LAG member as a monitor port in a port-mirroring configuration.
- LAGs cannot be monitored by a monitor port; however, LAG members can be monitored.
- All ports in static LAGs must have the same link configuration (speed, duplex, flow control).

Configuring a Static LAG

In the following example, three ports are aggregated between two switches.

Figure 10. LAG Configuration Example



Prior to configuring each switch in the preceding example, you must connect to the appropriate switch's Command Line Interface (CLI) as the administrator.

Note: For details about accessing and using any of the menu commands described in this example, see the Enterprise NOS *Command Reference*.

1. Connect the switch ports that will be members in the LAG.
2. Configure the LAG using these steps on the CN4093:
   a. Define a LAG.

```
CN 4093(config)# portchannel 1 port ext1,ext2,ext3 (Add ports to LAG 1)
CN 4093(config)# portchannel 1 enable
```

   b. Verify the configuration.

```
CN 4093(config)# show portchannel information
```

      Examine the resulting information. If any settings are incorrect, make appropriate changes.

3. Repeat the process on the other switch.

```
CN 4093(config)# portchannel 3 port 2,12,22
CN 4093(config)# portchannel 3 enable
```

   LAG 1 (on the CN4093) is now connected to LAG 3 on the Application Switch.

   Note: In this example, a CN4093 and an application switch are used. If a third-party device supporting link aggregation is used (such as Cisco routers and switches with EtherChannel technology or Sun's Quad Fast Ethernet Adapter), LAGs on the third-party device should be configured manually. Connection problems could arise when using automatic LAG negotiation on the third-party device.

4. Examine the aggregation information on each switch.

```
CN 4093# show portchannel information          (View aggregation information)
```

   Information about each port in each configured LAG is displayed. Make sure that LAGs consist of the expected ports and that each port is in the expected state.

   The following restrictions apply:

   - Any physical switch port can belong to only one LAG.
   - Up to 24 ports can belong to the same LAG.
   - All ports in static LAGs must be have the same link configuration (speed, duplex, flow control).
   - Aggregation from third-party devices must comply with Cisco® EtherChannel® technology.

Configurable LAG Hash Algorithm

Traffic in a LAG is statistically distributed among member ports using a *hash* process where various address and attribute bits from each transmitted frame are recombined to specify the particular LAG port the frame will use. The CN4093 uses the RTAG7 model for LAG hashing.

The switch can be configured to use a variety of hashing options. To achieve the most even traffic distribution, select options that exhibit a wide range of values for your particular network. Avoid hashing on information that is not usually present in the expected traffic or which does not vary.

The CN4093 supports the following hashing options, which can be used in any combination:

- For Layer 2 traffic, one of the following combinations may be applied:
  - Source MAC address (`smac`)

```
CN 4093(config)# portchannel thash l2thash l2-source-mac-address
```

  - Destination MAC address (`dmac`)

```
CN 4093(config)# portchannel thash l2thash l2-destination-mac-address
```

  - Both source and destination MAC address (enabled by default)

```
CN4093(config)# portchannel thash l2thash l2-source-destination-mac
```

Note: At least one Layer 2 option must always be enabled; The `smac` and `dmac` options may not both be disabled at the same time.

- For Layer 3 IPv4/IPv6 traffic, one of the following are permitted:
  - Source IP address (`sip`)

```
CN 4093(config)# portchannel thash l3thash l3-source-ip-address
```

  - Destination IP address (`dip`)

```
CN 4093(config)# portchannel thash l3thash l3-destination-ip-address
```

  - Both source and destination IP address (enabled by default)

```
CN4093(config)# portchannel thash l3thash l3-source-destination-ip
```

If Layer 2 hashing is preferred for Layer 3 traffic, disable the Layer 3 `sip` and `dip` hashing options and enable the `useL2` option:

```
CN 4093(config)# portchannel thash l3thash l3-use-l2-hash
```

Layer 3 traffic will then use Layer 2 options for hashing.

- Ingress port number (disabled by default)

```
CN 4093(config)# portchannel thash ingress
```

- Layer 4 port information (disabled by default)

```
CN 4093(config)# portchannel thash l4port
```

When enabled, Layer 4 port information (TCP, UPD, etc.) is added to the hash if available. The `L4port` option is ignored when Layer 4 information is not included in the packet (such as for Layer 2 packets) or when the `useL2` option is enabled.

Note: For MPLS packets, Layer 4 port information is excluded from the hash calculation. Instead, other IP fields are used, along with the first two MPLS labels.

The CN4093 supports the following FCoE hashing options:

```
CN 4093(config)# portchannel thash fcoe cntag-id
CN 4093(config)# portchannel thash fcoe destination-id
CN 4093(config)# portchannel thash fcoe fabric-id
CN 4093(config)# portchannel thash fcoe originator-id
CN 4093(config)# portchannel thash fcoe responder-id
CN 4093(config)# portchannel thash fcoe source-id
```

Link Aggregation Control Protocol

Link Aggregation Control Protocol (LACP) is an IEEE 802.3ad standard for grouping several physical ports into one logical port (known as a dynamic Link Aggregation Group) with any device that supports the standard. Please refer to IEEE 802.3ad-2002 for a full description of the standard.

IEEE 802.3ad allows standard Ethernet links to form a single Layer 2 link using the Link Aggregation Control Protocol (LACP). Link aggregation is a method of grouping physical link segments of the same media type and speed in full duplex, and treating them as if they were part of a single, logical link segment. If a link in a LACP LAG fails, traffic is reassigned dynamically to the remaining link or links of the dynamic LAG.

The CN4093 supports up to 24 ports in a single LACP LAG. It also supports a total of 52 LACP LAGs.

Note: LACP implementation in *Enterprise NOS* does not support the Churn machine, an option used to detect if the port is operable within a bounded time period between the actor and the partner. Only the Marker Responder is implemented and there is no marker protocol generator.

A port's Link Aggregation Identifier (LAG ID) determines how the port can be aggregated. The Link Aggregation ID (LAG ID) is constructed mainly from the *system ID* and the port's *admin key*, as follows:

- **System ID**: an integer value based on the switch's MAC address and the system priority assigned in the CLI.
- **Admin key:** a port's *admin key* is an integer value (1 - 65535) that you can configure in the CLI. Each CN4093 port that participates in the same LACP LAG must have the same *admin key* value. The admin key is *locally significant*, which means the partner switch does not need to use the same admin key value.

For example, consider two switches, an Actor (the CN4093) and a Partner (another switch), as shown in Table 16.

Table 16. *Actor vs. Partner LACP configuration*

Actor Switch	Partner Switch 1
Port 38 (admin key = 100)	Port 1 (admin key = 50)
Port 39 (admin key = 100)	Port 2 (admin key = 50)
Port 40 (admin key = 100)	Port 3 (admin key = 70)

In the configuration shown in Table 16, Actor switch ports 38 and 39 aggregate to form an LACP LAG with Partner switch ports 1 and 2. Only ports with the same LAG ID are aggregated in the LAG. Actor switch port 40 is not aggregated in the LAG because it has a different LAG ID. Switch ports configured with the same admin key on the Actor switch but have a different LAG ID (due to Partner switch admin key configuration or due to partner switch MAC address being different) can be aggregated in another LAG i.e. Actor switch port 40 can be aggregated in another LAG with ports that have the same LAG ID as port 40.

To avoid the Actor switch ports (with the same admin key) from aggregating in another LAG, you can configure a LAG ID. Ports with the same admin key (although with different LAG IDs) compete to get aggregated in a LAG. The LAG ID for the LAG is decided based on the first port that is aggregated in the LAG. Ports with this LAG ID get aggregated and the other ports are placed in `suspended` mode. As per the configuration shown in Table 16, if port 38 gets aggregated first, then the LAG ID of port 38 would be the LAG ID of the LAG. Port 40 would be placed in `suspended` mode. When in `suspended` mode, a port transmits only LACP data units (LACPDUs) and discards all other traffic.

A port may also be placed in suspended mode for the following reasons:

- When LACP is configured on the port but it stops receiving LACPDUs from the partner switch.
- When the port has a different LAG ID because of the partner switch MAC being different. For example: when a switch is connected to two partners.

LAG ID can be configured using the following command:

```
CN 4093(config)# portchannel <65-128> lacp key <adminkey of the LAG>
```

LACP provides for the controlled addition and removal of physical links for the link aggregation.

LACP Modes

Each port in the CN4093 can have one of the following LACP modes.

- `off` (default)

  The user can configure this port in to a regular static LAG.

- `active`

  The port is capable of forming a LACP LAG. This port sends LACPDU packets to partner system ports.

- `passive`

  The port is capable of forming a LACP LAG. This port only responds to the LACPDU packets sent from a LACP `active` port.

Each active LACP port transmits LACP data units (LACPDUs), while each passive LACP port listens for LACPDUs. During LACP negotiation, the admin key is exchanged. The LACP LAG is enabled as long as the information matches at both ends of the link. If the admin key value changes for a port at either end of the link, that port's association with the LACP LAG is lost.

If a LACP LAG member port is connected to a port that is in LACP `off` mode, the LACP port will not be able to converge and the link goes down.

When the system is initialized, all ports are by default in LACP `off` mode and are assigned unique *admin keys*. To make a group of ports capable of being aggregated, you assign them all the same *admin key*. You must set the port's LACP mode to `active` to activate LACP negotiation. You can set other port's LACP mode to passive, to reduce the amount of LACPDU traffic at the initial LAG-forming stage.

Use the following command to check whether the ports are aggregated.

```
CN 4093 # show lacp information
```

Static LAGs are listed as LAGs 1 though 52. Dynamic LAGs are listed as 65 through 128.

LACP individual

Ports assigned with an LACP admin key are prevented by default from forming individual links if they cannot join a LACP LAG. To override this behavior, use the following commands:

```
CN 4093(config)# interface port <port alias or number>
CN 4093(config-if)# no lacp suspend-individual
```

This allows the selected ports to be treated as normal link-up ports, which may forward data traffic according to STP, Hot Links or other applications, if they do not receive any LACPDUs.

To configure the LACP individual setting for all the ports in a static LACP LAG, use the following commands:

```
CN 4093(config-if)# interface portchannel lacp <LAG admin key>
CN 4093(config-if)# [no] lacp suspend-individual
```

Note: By default, ports are configured as below:

- external ports with **`lacp suspend-individual`**
- internal ports with **`no lacp suspend-individual`**

Configuring LACP

Use the following procedure to configure LACP for ports 7, 8 and 9 to participate in a single link aggregation.

1. Configure port parameters. All ports that participate in the LACP LAG must have the same settings, including VLAN membership.
2. Select the port range and define the admin key. Only ports with the same admin key can form a LACP LAG.

```
CN 4093(config)# interface port 7-9
CN 4093(config-if)# lacp key 100
```

3. Set the LACP mode.

```
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# exit
```

4. Optionally allow member ports to individually participate in normal data traffic if no LACPDUs are received.

```
CN 4093(config-if)# no lacp suspend-individual
CN 4093(config-if)# exit
```

5. Set the link aggregation as static, by associating it with LAG ID 65:

```
CN 4093(config-if)# portchannel 65 lacp key 100
```

Chapter 10. Spanning Tree Protocols

When multiple paths exist between two points on a network, Spanning Tree Protocol (STP), or one of its enhanced variants, can prevent broadcast loops and ensure that the CN4093 10 Gb Converged Scalable Switch (CN4093) uses only the most efficient network path.

This chapter covers the following topics:

Spanning Tree Protocol Modes

Enterprise NOS 8.4 supports the following STP modes:

- Rapid Spanning Tree Protocol (RSTP)

  IEEE 802.1D (2004) RSTP allows devices to detect and eliminate logical loops in a bridged or switched network. When multiple paths exist, STP configures the network so that only the most efficient path is used. If that path fails, STP automatically configures the best alternative active path on the network in order to sustain network operations. RSTP is an enhanced version of IEEE 802.1D (1998) STP, providing more rapid convergence of the Spanning Tree network path states on STG 1.

  See "Rapid Spanning Tree Protocol" on page 189 for details.

- Per-VLAN Rapid Spanning Tree (PVRST+)

  PVRST mode is based on RSTP to provide rapid Spanning Tree convergence, but supports instances of Spanning Tree, allowing one STG per VLAN. PVRST mode is compatible with Cisco R-PVST/R-PVST+ mode.

  PVRST is the default Spanning Tree mode on the CN4093. See "PVRST Mode" on page 177 for details.

- Multiple Spanning Tree Protocol (MSTP)

  IEEE 802.1Q (2003) MSTP provides both rapid convergence and load balancing in a VLAN environment. MSTP allows multiple STGs, with multiple VLANs in each.

  See "Multiple Spanning Tree Protocol" on page 191 for details.

Global STP Control

By default, the Spanning Tree feature is globally enabled on the switch, and is set for PVRST mode. Spanning Tree (and thus any currently configured STP mode) can be globally disabled or re-enabled using the following commands:

```
CN 4093(config)# spanning-tree mode disable      (Globally disable Spanning Tree)
```

Spanning Tree can be re-enabled by specifying the STP mode:

```
CN 4093(config)# spanning-tree mode {pvrst|rstp|mst}
```

PVRST Mode

Note: Per-VLAN Rapid Spanning Tree (PVRST) is enabled by default on the CN4093.

Using STP, network devices detect and eliminate logical loops in a bridged or switched network. When multiple paths exist, Spanning Tree configures the network so that a switch uses only the most efficient path. If that path fails, Spanning Tree automatically sets up another active path on the network to sustain network operations.

Enterprise NOS PVRST mode is based on IEEE 802.1w RSTP. Like RSTP, PVRST mode provides rapid Spanning Tree convergence. However, PVRST mode is enhanced for multiple instances of Spanning Tree. In PVRST mode, each VLAN may be automatically or manually assigned to one of 127 available STGs, with each STG acting as an independent, simultaneous instance of STP. PVRST uses IEEE 802.1Q tagging to differentiate STP BPDUs and is compatible with Cisco R-PVST/R-PVST+ modes.

The relationship between ports, LAGs, VLANs and Spanning Trees is shown in Table 17.

Table 17. *Ports, LAGs and VLANs*

Switch Element	Belongs To
Port	LAG or one or more VLANs
LAG	One or more VLANs
VLAN (non-default)	• PVRST: One VLAN per STG • RSTP: All VLANs are in STG 1 • MSTP: Multiple VLANs per STG

Port States

The port state controls the forwarding and learning processes of Spanning Tree. In PVRST, the port state has been consolidated to the following: `discarding`, `learning` or `forwarding`.

Due to the sequence involved in these STP states, considerable delays may occur while paths are being resolved. To mitigate delays, ports defined as *edge* ports ("Port Type and Link Type" on page 195) may bypass the `discarding` and `learning` states, and enter directly into the `forwarding` state.

Bridge Protocol Data Units

To create a Spanning Tree, the switch generates a configuration Bridge Protocol Data Unit (BPDU), which it then forwards out of its ports. All switches in the Layer 2 network participating in the Spanning Tree gather information about other switches in the network through an exchange of BPDUs.

A bridge sends BPDU packets at a configurable regular interval (2 seconds by default). The BPDU is used to establish a path, much like a hello packet in IP routing. BPDUs contain information about the transmitting bridge and its ports, including bridge MAC addresses, bridge priority, port priority, and path cost. If the ports are tagged, each port sends out a special BPDU containing the tagged information.

The generic action of a switch on receiving a BPDU is to compare the received BPDU to its own BPDU that it will transmit. If the priority of the received BPDU is better than its own priority, it will replace its BPDU with the received BPDU. Then, the switch adds its own bridge ID number and increments the path cost of the BPDU. The switch uses this information to block any necessary ports.

Note: If STP is globally disabled, BPDUs from external devices will transit the switch transparently. If STP is globally enabled, for ports where STP is turned off, inbound BPDUs will instead be discarded.

Determining the Path for Forwarding BPDUs

When determining which port to use for forwarding and which port to block, the CN4093 uses information in the BPDU, including each bridge ID. A technique based on the "lowest root cost" is then computed to determine the most efficient path for forwarding.

Bridge Priority

The bridge priority parameter controls which bridge on the network is the STG root bridge. To make one switch become the root bridge, configure the bridge priority lower than all other switches and bridges on your network. The lower the value, the higher the bridge priority. Use the following command to configure the bridge priority:

```
CN 4093(config)# spanning-tree stp <1-128> bridge priority <0-65535>
```

Port Priority

The port priority helps determine which bridge port becomes the root port or the designated port. The case for the root port is when two switches are connected using a minimum of two links with the same path-cost. The case for the designated port is in a network topology that has multiple bridge ports with the same path-cost connected to a single segment, the port with the lowest port priority becomes the designated port for the segment.

Use the following commands to configure the port priority:

```
CN 4093(config)# interface port <port alias or number>
CN 4093(config-if)# spanning-tree stp <1-128> priority <0-65535>
```

where *priority value* is a number from 0 to 240, in increments of 16 (such as 0, 16, 32, and so on). If the specified priority value is not evenly divisible by 16, the value will be automatically rounded down to the nearest valid increment whenever manually changed in the configuration.

Root Guard

The root guard feature provides a way to enforce the root bridge placement in the network. It keeps a new device from becoming root and thereby forcing STP re-convergence. If a root-guard enabled port detects a root device, that port will be placed in a blocked state.

You can configure the root guard at the port level using the following commands:

```
CN 4093(config)# interface port <port alias or number>
CN 4093(config-if)# spanning-tree guard root
```

The default state is none (disabled).

Loop Guard

In general, STP resolves redundant network topologies into loop-free topologies. The loop guard feature performs additional checking to detect loops that might not be found using Spanning Tree. STP loop guard ensures that a non-designated port does not become a designated port.

To globally enable loop guard, enter the following command:

```
CN 4093(config)# spanning-tree loopguard
```

Note: The global loop guard command will be effective on a port only if the port-level loop guard command is set to default as shown below:
`CN 4093(config-if)# spanning-tree guard loop none`

To enable loop guard at the port level, enter the following command:

```
CN 4093(config)# interface port <port alias or number>
CN 4093(config-if)# spanning-tree guard loop
```

The default state is “none”, i.e. disabled.

Port Path Cost

The port path cost assigns lower values to high-bandwidth ports, such as 10 Gigabit Ethernet, to encourage their use. The cost of a port also depends on whether the port operates at full-duplex (lower cost) or half-duplex (higher cost). For example, if a 100-Mbps (Fast Ethernet) link has a "cost" of 10 in half-duplex mode, it will have a cost of 5 in full-duplex mode. The objective is to use the fastest links so that the route with the lowest cost is chosen. A value of 0 (the default) indicates that the default cost will be computed for an auto-negotiated link or LAG speed.

Use the following command to modify the port path cost:

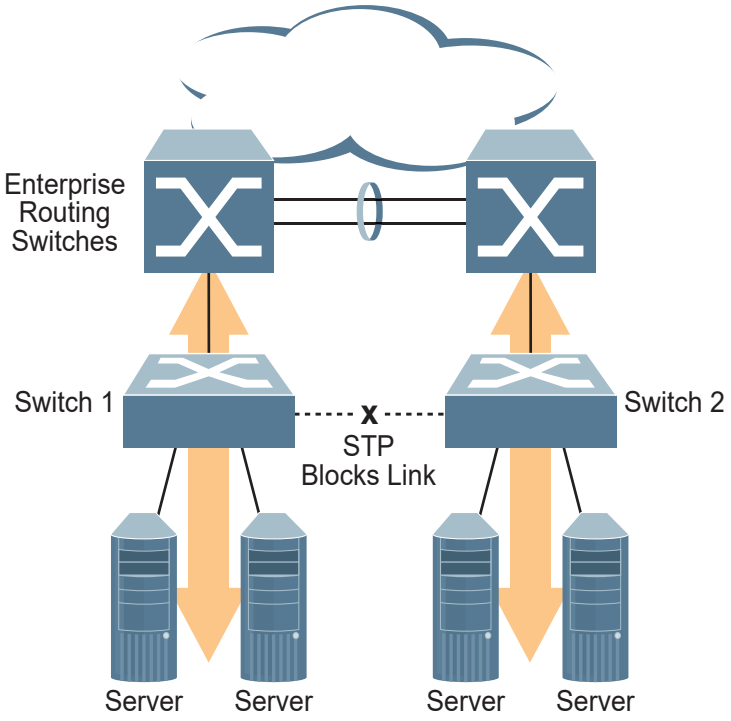
```
CN 4093(config)# interface port <port alias or number>
CN 4093(config-if)# spanning-tree stp <1-128> path-cost <path cost value>
CN 4093(config-if)# exit
```

The port path cost can be a value from 1 to 200000000. Specify 0 for automatic path cost.

Simple STP Configuration

Figure 11 depicts a simple topology using a switch-to-switch link between two switches (via either external ports or internal Inter-Switch Links).

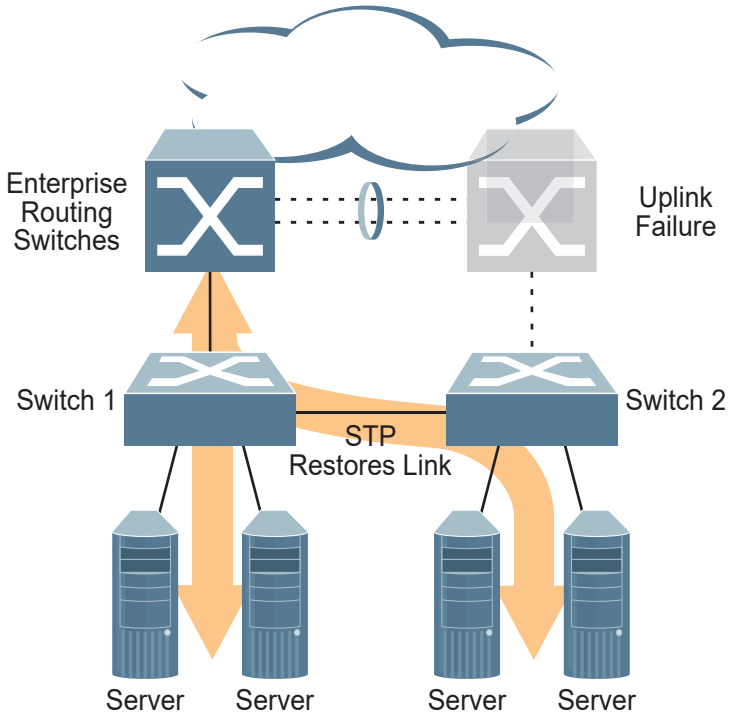
Figure 11. Spanning Tree Blocking a Switch-to-Switch Link



To prevent a network loop among the switches, STP must block one of the links between them. In this case, it is desired that STP block the link between the blade switches, and not one of the CN4093 uplinks or the Enterprise switch LAG.

During operation, if one CN4093 experiences an uplink failure, STP will activate the switch-to-switch link so that server traffic on the affected CN4093 may pass through to the active uplink on the other CN4093, as shown in Figure 12.

Figure 12. Spanning Tree Restoring the Switch-to-Switch Link



In this example, port 10 on each switch is used for the switch-to-switch link. To ensure that the CN4093 switch-to-switch link is blocked during normal operation, the port path cost is set to a higher value than other paths in the network. To configure the port path cost on the switch-to-switch links in this example, use the following commands on each switch.

```
CN 4093(config)# interface port 10
CN 4093(config-if)# spanning-tree stp 1 path-cost 60000
CN 4093(config-if)# exit
```

Per-VLAN Spanning Tree Groups

PVRST mode supports a maximum of 128 STGs, with each STG acting as an independent, simultaneous instance of STP.

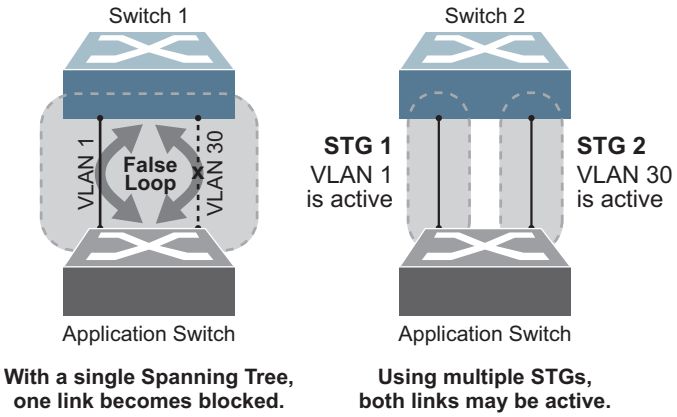
Multiple STGs provide multiple data paths which can be used for load-balancing and redundancy. To enable load balancing between two CN4093s using multiple STGs, configure each path with a different VLAN and then assign each VLAN to a separate STG. Since each STG is independent, they each send their own IEEE 802.1Q tagged Bridge Protocol Data Units (BPDUs).

Each STG behaves as a bridge group and forms a loop-free topology. The default STG 1 may contain multiple VLANs (typically until they can be assigned to another STG). STGs 2-128 may contain only one VLAN each.

Using Multiple STGs to Eliminate False Loops

Figure 13 shows a simple example of why multiple STGs are needed. In the figure, two ports on a CN4093 are connected to two ports on an application switch. Each of the links is configured for a different VLAN, preventing a network loop. However, in the first network, since a single instance of Spanning Tree is running on all the ports of the CN4093, a physical loop is assumed to exist, and one of the VLANs is blocked, impacting connectivity even though no actual loop exists.

Figure 13. Using Multiple Instances of Spanning Tree Group



In the second network, the problem of improper link blocking is resolved when the VLANs are placed into different Spanning Tree Groups (STGs). Since each STG has its own independent instance of Spanning Tree, each STG is responsible only for the loops within its own VLAN. This eliminates the false loop, and allows both VLANs to forward packets between the switches at the same time.

VLAN and STG Assignment

In PVRST mode, up to 128 STGs are supported. Ports cannot be added directly to an STG. Instead, ports must be added as members of a VLAN, and the VLAN must then be assigned to the STG.

STG 1 is the default STG. Although VLANs can be added to or deleted from default STG 1, the STG itself cannot be deleted from the system. By default, STG 1 is enabled and includes VLAN 1, which by default includes all switch ports (except for management VLANs and management ports).

STG 128 is reserved for switch management. By default, STG 128 is disabled, but includes management VLAN 4095 and the management ports.

By default, all other STGs (STG 2 through 127) are enabled, though they initially include no member VLANs. VLANs must be assigned to STGs. By default, this is done automatically using VLAN Automatic STG Assignment (VASA), though it can also be done manually (see "Manually Assigning STGs" on page 184).

When VASA is enabled (as by default), each time a new VLAN is configured, the switch will automatically assign that new VLAN to its own STG. Conversely, when a VLAN is deleted, if its STG is not associated with any other VLAN, the STG is returned to the available pool.

The specific STG number to which the VLAN is assigned is based on the VLAN number itself. For low VLAN numbers (1 through 127), the switch will attempt to assign the VLAN to its matching STG number. For higher numbered VLANs, the STG assignment is based on a simple modulus calculation; the attempted STG number will "wrap around," starting back at the top of STG list each time the end of the list is reached. However, if the attempted STG is already in use, the switch will select the next available STG. If an empty STG is not available when creating a new VLAN, the VLAN is automatically assigned to default STG 1.

If ports are tagged, each tagged port sends out a special BPDU containing the tagged information. Also, when a tagged port belongs to more than one STG, the egress BPDUs are tagged to distinguish the BPDUs of one STG from those of another STG.

VASA is enabled by default, but can be disabled or re-enabled using the following command:

```
CN 4093(config)# [no] spanning-tree stg-auto
```

If VASA is disabled, when you create a new VLAN, that VLAN automatically belongs to default STG 1. To place the VLAN in a different STG, assign it manually.

VASA applies only to PVRST mode and is ignored in RSTP and MSTP modes.

Manually Assigning STGs

The administrator may manually assign VLANs to specific STGs, whether or not VASA is enabled.

1. If no VLANs exist (other than default VLAN 1), see "Guidelines for Creating VLANs" on page 184 for information about creating VLANs and assigning ports to them.
2. Assign the VLAN to an STG using one of the following methods:
   - From the global configuration mode:

```
CN 4093(config)# spanning-tree stp <1-128> vlan <VLAN>
```

   - Or from within the VLAN configuration mode:

```
CN 4093(config)# vlan <VLAN number>
CN 4093(config-vlan)# stg <STG number>
CN 4093(config-vlan)# exit
```

When a VLAN is assigned to a new STG, the VLAN is automatically removed from its prior STG.

Note: For proper operation with switches that use Cisco PVST+, it is recommended that you create a separate STG for each VLAN.

Guidelines for Creating VLANs

- When you create a new VLAN, if VASA is enabled (the default), that VLAN is automatically assigned its own STG. If VASA is disabled, the VLAN automatically belongs to STG 1, the default STG. To place the VLAN in a different STG, see "Manually Assigning STGs" on page 184. The VLAN is automatically removed from its old STG before being placed into the new STG.
- Each VLANs must be contained *within* a single STG; a VLAN cannot span multiple STGs. By confining VLANs within a single STG, you avoid problems with Spanning Tree blocking ports and causing a loss of connectivity within the VLAN. When a VLAN spans multiple switches, it is recommended that the VLAN remain within the same STG (be assigned the same STG ID) across all the switches.
- If ports are tagged, all aggregated ports can belong to multiple STGs.
- A port cannot be directly added to an STG. The port must first be added to a VLAN, and that VLAN added to the desired STG.

Rules for VLAN Tagged Ports

- Tagged ports can belong to more than one STG, but untagged ports can belong to only one STG.
- When a tagged port belongs to more than one STG, the egress BPDUs are tagged to distinguish the BPDUs of one STG from those of another STG.

Adding and Removing Ports from STGs

- When you add a port to a VLAN that belongs to an STG, the port is also added to that STG. However, if the port you are adding is an untagged port and is already a member of another STG, that port will be removed from its current STG and added to the new STG. An untagged port cannot belong to more that one STG.

  For example: Assume that VLAN 1 belongs to STG 1, and that port 1 is untagged and does not belong to any STG. When you add port 1 to VLAN 1, port 1 will automatically become part of STG 1.

  However, if port 5 is untagged and is a member of VLAN 3 in STG 2, then adding port 5 to VLAN 1 in STG 1 will not automatically add the port to STG 1. Instead, the switch will prompt you to decide whether to change the PVID from 3 to 1:

  ```
  "Port 5 is an UNTAGGED/Access Mode port and its current PVID/Native
  VLAN is 3.
  Confirm changing PVID/Native VLAn from 3 to 1 [y/n]:" y
  ```

- When you remove a port from VLAN that belongs to an STG, that port will also be removed from the STG. However, if that port belongs to another VLAN in the same STG, the port remains in the STG.

  As an example, assume that port 2 belongs to only VLAN 2, and that VLAN 2 belongs to STG 2. When you remove port 2 from VLAN 2, the port is moved to default VLAN 1 and is removed from STG 2.

  However, if port 2 belongs to both VLAN 1 and VLAN 2, and both VLANs belong to STG 1, removing port 2 from VLAN 2 does not remove port 2 from STG 1, because the port is still a member of VLAN 1, which is still a member of STG 1.

- An STG cannot be deleted, only disabled. If you disable the STG while it still contains VLAN members, Spanning Tree will be off on all ports belonging to that VLAN.

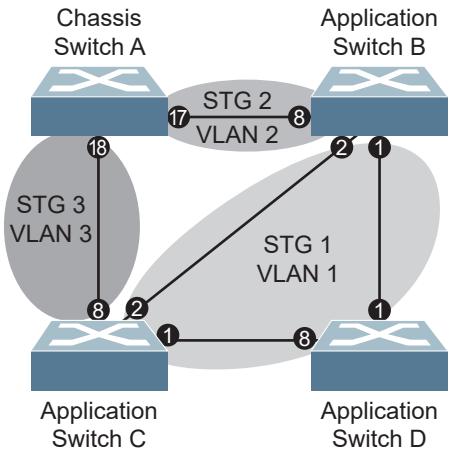
The relationship between port, LAGs, VLANs and Spanning Trees is shown in Table 17 on page 177.

Switch-Centric Configuration

PVRST is switch-centric: STGs are enforced only on the switch where they are configured. The STG ID is not transmitted in the Spanning Tree BPDU. Each Spanning Tree decision is based entirely on the configuration of the particular switch.

For example, in Figure 14, though VLAN 2 is shared by the Switch A and Switch B, each switch is responsible for the proper configuration of its own ports, VLANs, and STGs. Switch A identifies its own port 17 as part of VLAN 2 on STG 2, and the Switch B identifies its own port 8 as part of VLAN 2 on STG 2.

Figure 14. Implementing Multiple Spanning Tree Groups



The VLAN participation for each Spanning Tree Group in Figure 14 on page 186 is as follows:

- VLAN 1 Participation

  Assuming Switch B to be the root bridge, Switch B transmits the BPDU for VLAN 1 on ports 1 and 2. Switch C receives the BPDU on port 2, and Switch D receives the BPDU on port 1. Because there is a network loop between the switches in VLAN 1, either Switch D will block port 8 or Switch C will block port 1, depending on the information provided in the BPDU.

- VLAN 2 Participation

  Switch B, the root bridge, generates a BPDU for STG 2 from port 8. Switch A receives this BPDU on port 17, which is assigned to VLAN 2, STG 2. Because switch B has no additional ports participating in STG 2, this BPDU is not forwarded to any additional ports and Switch B remains the designated root.

- VLAN 3 Participation

  For VLAN 3, Switch A or Switch C may be the root bridge. If Switch A is the root bridge for VLAN 3, STG 3, then Switch A transmits the BPDU from port 18. Switch C receives this BPDU on port 8 and is identified as participating in VLAN 3, STG 3. Since Switch C has no additional ports participating in STG 3, this BPDU is not forwarded to any additional ports and Switch A remains the designated root.

Configuring Multiple STGs

This configuration shows how to configure the three instances of STGs on the switches A, B, C and D illustrated in Figure 14 on page 186.

Because VASA is enabled by default, each new VLAN is automatically assigned its own STG. However, for this configuration example, some VLANs are explicitly reassigned to other STGs.

1. Set the Spanning Tree mode on each switch to PVRST.

```
CN 4093(config)# spanning-tree mode pvrst
```

Note: PVRST is the default mode on the CN4093. This step is not required unless the STP mode has been previously changed, and is shown here merely as an example of manual configuration.

2. Configure the following on Switch A:

   a. Enable VLAN 2 and VLAN 3.

```
CN 4093(config)# vlan 2
CN 4093(config-vlan)# exit
CN 4093(config)# vlan 3
CN 4093(config-vlan)# exit

If VASA is disabled, enter the following commands:
CN 4093(config)# spanning-tree stp 2 vlan 2
CN 4093(config)# spanning-tree stp 3 vlan 3
```

   b. Add port 17 to VLAN 2 and port 18 to VLAN 3.

```
CN 4093(config)# interface port 17
CN 4093(config-if)# switchport access vlan 2
CN 4093(config-if)# exit

CN 4093(config)# interface port 18
CN 4093(config-if)# switchport access vlan 3
CN 4093(config-if)# exit
```

   VLAN 2 and VLAN 3 are removed from STG 1.

   Note: In PVRST mode, each instance of STG is enabled by default.

3. Configure the following on Switch B:

   Add port 8 to VLAN 2. Ports 1 and 2 are by default in VLAN 1 assigned to STG 1.

```
CN 4093(config)# vlan 2
CN 4093(config-vlan)# stg 2
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 8
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit

If VASA is disabled, enter the following command:
CN 4093(config)# spanning-tree stp 2 vlan 2
```

VLAN 2 is automatically removed from STG 1. By default VLAN 1 remains in STG 1.

4. Configure the following on application switch C:

   Add port 8 to VLAN 3. Ports 1 and 2 are by default in VLAN 1 assigned to STG 1.

```
CN 4093(config)# vlan 3
CN 4093(config-vlan)# stg 3
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 8
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit

If VASA is disabled, enter the following command:
CN 4093(config)# spanning-tree stp 3 vlan 3
```

   VLAN 3 is automatically removed from STG 1. By default VLAN 1 remains in STG 1.

   Switch D does not require any special configuration for multiple Spanning Trees. Switch D uses default STG 1 only.

Rapid Spanning Tree Protocol

RSTP provides rapid convergence of the Spanning Tree and provides the fast re-configuration critical for networks carrying delay-sensitive traffic such as voice and video. RSTP significantly reduces the time to reconfigure the active topology of the network when changes occur to the physical topology or its configuration parameters. RSTP reduces the bridged-LAN topology to a single Spanning Tree.

RSTP was originally defined in IEEE 802.1w (2001) and was later incorporated into IEEE 802.1D (2004), superseding the original STP standard.

RSTP parameters apply only to Spanning Tree Group (STG) 1. The PVRST mode STGs 2-128 are not used when the switch is placed in RSTP mode.RSTP is compatible with devices that run IEEE 802.1D (1998) Spanning Tree Protocol. If the switch detects IEEE 802.1D (1998) BPDUs, it responds with IEEE 802.1D (1998)-compatible data units. RSTP is not compatible with Per-VLAN Rapid Spanning Tree (PVRST) protocol.

Note: In RSTP mode, Spanning Tree for the management ports is turned off by default.

Port States

RSTP port state controls are the same as for PVRST: `discarding`, `learning` and `forwarding`.

Due to the sequence involved in these STP states, considerable delays may occur while paths are being resolved. To mitigate delays, ports defined as *edge* ports ("Port Type and Link Type" on page 195) may bypass the `discarding` and `learning` states, and enter directly into the `forwarding` state.

RSTP Configuration Guidelines

This section provides important information about configuring RSTP. When RSTP is turned on, the following occurs:

- STP parameters apply only to STG 1.
- Only STG 1 is available. All other STGs are turned off.
- All VLANs, including management VLANs, are moved to STG 1.

RSTP Configuration Example

This section provides steps to configure RSTP.

1. Configure port and VLAN membership on the switch.
2. Set the Spanning Tree mode to Rapid Spanning Tree.

```
CN 4093(config)# spanning-tree mode rstp
```

3. Configure RSTP parameters.

```
CN 4093(config)# spanning-tree stp 1 bridge priority 8192
CN 4093(config)# spanning-tree stp 1 bridge hello-time 5
CN 4093(config)# spanning-tree stp 1 bridge forward-delay 20
CN 4093(config)# spanning-tree stp 1 bridge maximum-age 30
CN 4093(config)# no spanning-tree stp 1 enable
```

4. Configure port parameters:

```
CN 4093(config)# interface port 3
CN 4093(config-if)# spanning-tree stp 1 priority 240
CN 4093(config-if)# spanning-tree stp 1 path-cost 500
CN 4093(config-if)# no spanning-tree stp 1 enable
CN 4093(config-if)# exit
```

Multiple Spanning Tree Protocol

Multiple Spanning Tree Protocol (MSTP) extends Rapid Spanning Tree Protocol (RSTP), allowing multiple Spanning Tree Groups (STGs) which may each include multiple VLANs. MSTP was originally defined in IEEE 802.1s (2002) and was later included in IEEE 802.1Q (2003).

In MSTP mode, the CN4093 supports up to 32 instances of Spanning Tree, corresponding to STGs 1-32, with each STG acting as an independent, simultaneous instance of STP.

MSTP allows frames assigned to different VLANs to follow separate paths, with each path based on an independent Spanning Tree instance. This approach provides multiple forwarding paths for data traffic, thereby enabling load-balancing, and reducing the number of Spanning Tree instances required to support a large number of VLANs.

Due to Spanning Tree's sequence of discarding, learning, and forwarding, lengthy delays may occur while paths are being resolved. Ports defined as *edge* ports ("Port Type and Link Type" on page 195) bypass the Discarding and Learning states, and enter directly into the Forwarding state.

Note: In MSTP mode, Spanning Tree for the management ports is turned off by default.

MSTP Region

A group of interconnected bridges that share the same attributes is called an MST region. Each bridge within the region must share the following attributes:

- Alphanumeric name
- Revision number
- VLAN-to STG mapping scheme

MSTP provides rapid re-configuration, scalability and control due to the support of regions, and multiple Spanning-Tree instances support within each region.

Common Internal Spanning Tree

The Common Internal Spanning Tree (CIST) provides a common form of Spanning Tree Protocol, with one Spanning-Tree instance that can be used throughout the MSTP region. CIST allows the switch to interoperate with legacy equipment, including devices that run IEEE 802.1D (1998) STP.

CIST allows the MSTP region to act as a virtual bridge to other bridges outside of the region, and provides a single Spanning-Tree instance to interact with them.

CIST port configuration includes Hello time, Edge port enable/disable, and Link Type. These parameters do not affect Spanning Tree Groups 1–32. They apply only when the CIST is used.

MSTP Configuration Guidelines

This section provides important information about configuring Multiple Spanning Tree Groups:

- When MSTP is turned on, the switch automatically moves management VLAN 4095 to the CIST. When MSTP is turned off, the switch moves VLAN 4095 from the CIST to Spanning Tree Group 128.
- When you enable MSTP, you must configure the Region Name. A default version number of 0 is configured automatically.
- Each bridge in the region must have the same name, version number and VLAN mapping.

MSTP Configuration Examples

MSTP Configuration Example 1

This section provides steps to configure MSTP on the CN4093.

1. Configure port and VLAN membership on the switch.
2. Configure Multiple Spanning Tree region parameters and set the mode to MSTP.

```
CN 4093(config)# spanning-tree mst configuration (Enter MST configuration mode)
CN 4093(config-mst)# name <name>                  (Define the Region name)
CN 4093(config-mst)# revision 100                 (Define the Revision level)
CN 4093(config-mst)# exit
CN 4093(config)# spanning-tree mode mst       (Set mode to Multiple Spanning Trees)
```

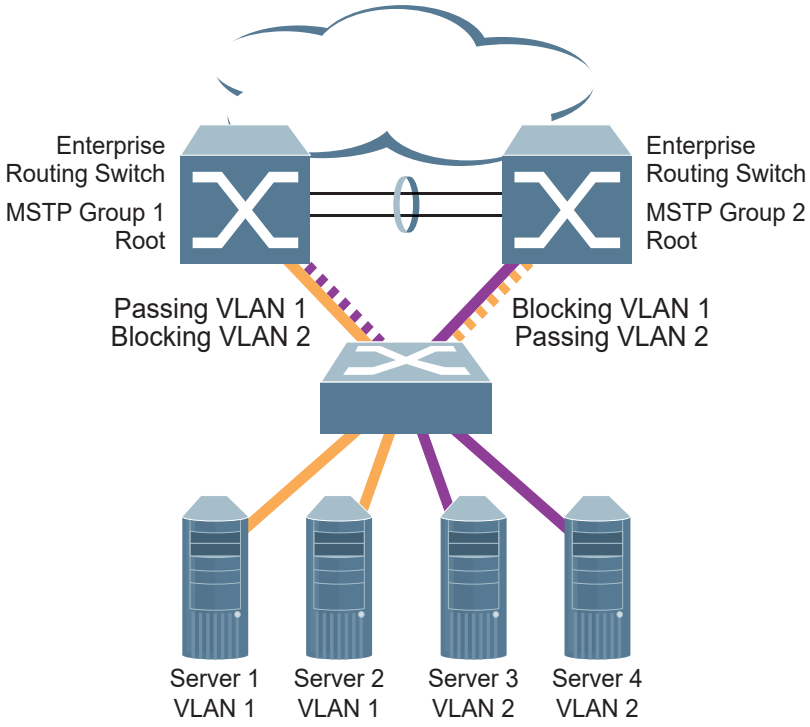
3. Map VLANs to MSTP instances:

```
CN 4093(config)# spanning-tree mst configuration (Enter MST configuration mode)
CN 4093(config-mst)# instance <1-32> vlan <vlan number or range>
```

MSTP Configuration Example 2

This configuration shows how to configure MSTP Groups on the switch, as shown in Figure 15.

Figure 15. Implementing Multiple Spanning Tree Groups



This example shows how multiple Spanning Trees can provide redundancy without wasting any uplink ports. In this example, the server ports are split between two separate VLANs. Both VLANs belong to two different MSTP groups. The Spanning Tree *priority* values are configured so that each routing switch is the root for a different MSTP instance. All of the uplinks are active, with each uplink port backing up the other.

1. Configure port membership and define the STGs for VLAN 1. Enable tagging on uplink ports that share VLANs. Port 19 and port 20 connect to the Enterprise Routing switches.

```
CN 4093(config)# interface port 19,20
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit
```

2. Configure MSTP: Spanning Tree mode, region name, and version.

```
CN 4093(config)# spanning-tree mst configuration
CN 4093(config-mst)# name MyRegion      (Define the Region name)
CN 4093(config-mst)# revision 100       (Define the Revision level)
CN 4093(config-mst)# exit
CN 4093(config)# spanning-tree mode mst     (Set mode to Multiple Spanning Trees)
```

3. Map VLANs to MSTP instances:

```
CN 4093(config)# spanning-tree mst configuration
CN 4093(config-mst)# instance 1 vlan 1
CN 4093(config-mst)# instance 2 vlan 2
```

4. Configure port membership and define the STGs for VLAN 2. Add server ports 3, 4 and 5 to VLAN 2. Add uplink ports 19 and 20 to VLAN 2. Assign VLAN 2 to STG 2.

```
CN 4093(config)# interface port 3,4,5,19,20
CN 4093(config-if)# switchport access vlan 2
CN 4093(config-if)# exit
```

Note: Each STG is enabled by default.

Port Type and Link Type

Edge/Portfast Port

A port that does not connect to a bridge is called an *edge port*. Since edge ports are assumed to be connected to non-STP devices (such as directly to hosts or servers), they are placed in the forwarding state as soon as the link is up. Internal ports (INT*x*) should be configured as edge ports.

Edge ports send BPDUs to upstream STP devices like normal STP ports, but should not receive BPDUs. If a port with `edge` enabled does receive a BPDU, it immediately begins working as a normal (non-edge) port, and participates fully in Spanning Tree.

Use the following commands to define or clear a port as an edge port:

```
CN 4093(config)# interface port <port>
CN 4093(config-if)# [no] spanning-tree portfast
CN 4093(config-if)# exit
```

Link Type

The link type determines how the port behaves in regard to Rapid Spanning Tree. Use the following commands to define the link type for the port:

```
CN 4093(config)# interface port <port>
CN 4093(config-if)# [no] spanning-tree link-type <type>
CN 4093(config-if)# exit
```

where *type* corresponds to the duplex mode of the port, as follows:

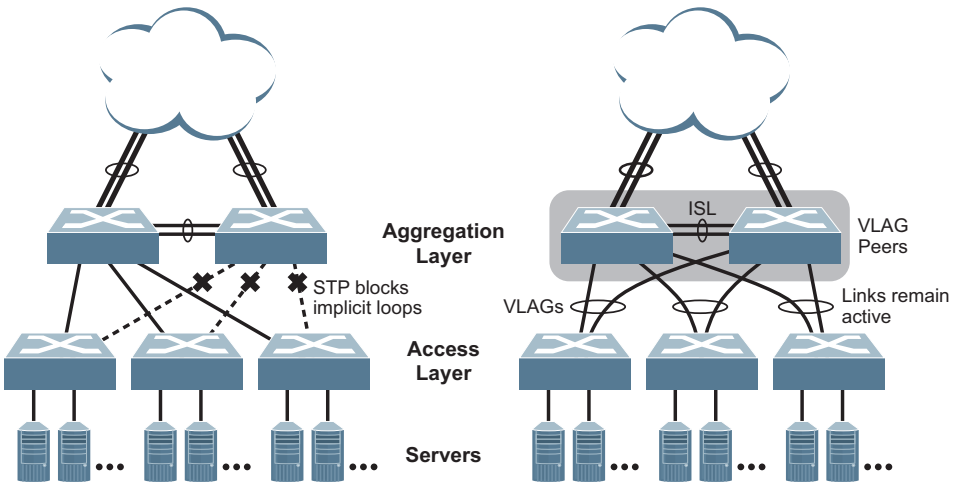
- `p2p` A full-duplex link to another device (point-to-point)
- `shared` A half-duplex link is a shared segment and can contain more than one device.
- `auto` The switch dynamically configures the link type.

Note: Any STP port in full-duplex mode can be manually configured as a shared port when connected to a non-STP-aware shared device (such as a typical Layer 2 switch) used to interconnect multiple STP-aware devices.

Chapter 11. Virtual Link Aggregation Groups

In many data center environments, downstream servers or switches connect to upstream devices which consolidate traffic. For example, see Figure 16.

Figure 16. Typical Data Center Switching Layers with STP vs. VLAG



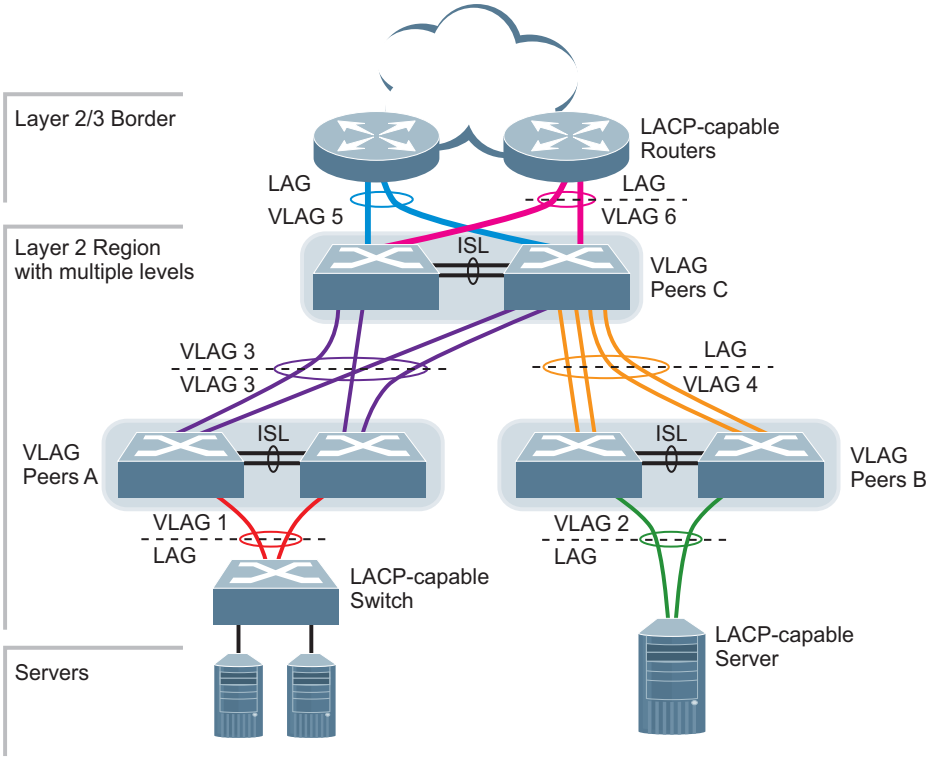
As shown in the example, a switch in the access layer may be connected to more than one switch in the aggregation layer to provide for network redundancy. Typically, Spanning Tree Protocol (RSTP, PVRST, or MSTP—see "Spanning Tree Protocols" on page 175) is used to prevent broadcast loops, blocking redundant uplink paths. This has the unwanted consequence of reducing the available bandwidth between the layers by as much as 50%. In addition, STP may be slow to resolve topology changes that occur during a link failure, and can result in considerable MAC address flooding.

Using Virtual Link Aggregation Groups (VLAGs), the redundant uplinks remain active, utilizing all available bandwidth.

Two switches are paired into VLAG peers, and act as a single virtual entity for the purpose of establishing a multi-port aggregation. Ports from both peers can be grouped into a VLAG and connected to the same LAG-capable target device. From the perspective of the target device, the ports connected to the VLAG peers appear to be a single LAG connecting to a single logical device. The target device uses the configured Tier ID to identify the VLAG peers as this single logical device. It is important that you use a unique Tier ID for each VLAG pair you configure. The VLAG-capable switches synchronize their logical view of the access layer port structure and internally prevent implicit loops. The VLAG topology also responds more quickly to link failure and does not result in unnecessary MAC flooding.

VLAGs are also useful in multi-layer environments for both uplink and downlink redundancy to any regular LAG-capable device. For example:

Figure 17. VLAG Application with Multiple Layers

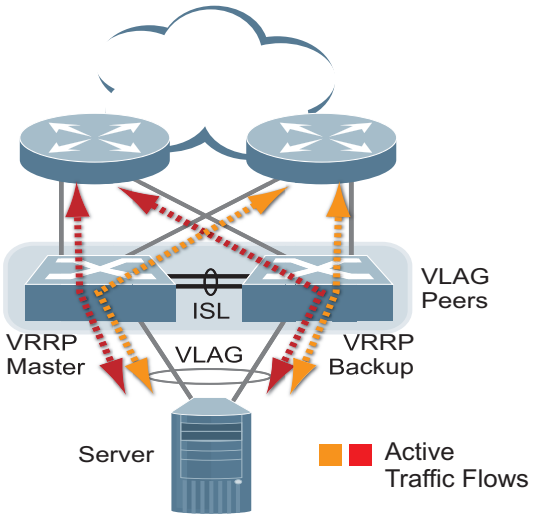


Wherever ports from *both* peered switches are aggregated to another device, the aggregated ports must be configured as a VLAG. For example, VLAGs 1 and 3 must be configured for both VLAG Peer A switches. VLAGs 2 and 4 must be configured for both VLAG Peer B switches.VLAGs 3, 5 and 6 must be configured on both VLAG Peer C switches. Other devices connecting to the VLAG peers are configured using regular static or dynamic LAGs.

Note: Do not configure a VLAG for connecting only one switch in the peer set to another device or peer set. For instance, in VLAG Peer C, a regular LAG is employed for the downlink connection to VLAG Peer B because only one of the VLAG Peer C switches is involved.

In addition, when used with VRRP, VLAGs can provide seamless active-active failover for network links. For example:

Figure 18. VLAG Application with VRRP



Note: VLAG is not compatible with UFP vPorts on the same ports.

VLAG Capacities

Servers or switches that connect to the VLAG peers using a multi-port VLAG are considered VLAG clients. VLAG clients are not required to be VLAG-capable. The ports participating in the VLAG are configured as regular port LAGs on the VLAG client end.

On the VLAG peers, the VLAGs are configured similarly to regular port LAGs, using many of the same features and rules. See "Ports and Link Aggregation (LAG)" on page 161 for general information concerning all port LAGs.

Each VLAG begins as a regular port LAG on each VLAG-peer switch. The VLAG may be either a static LAG (portchannel) or dynamic LACP LAG and consumes one slot from the overall port LAG capacity pool. The type of aggregation must match that used on VLAG client devices. Additional configuration is then required to implement the VLAG on both VLAG peer switches.

You may configure up to 52 LAGs on the switch, with all types (regular or VLAG, static or LACP) sharing the same pool.

The maximum number of configurable VLAG instances is as follows:

- **With STP off:** Maximum of 31 VLAG instances
- **With STP on:**
  - **PVRST/MSTP with one VLAG instance per VLAN/STG:** Maximum of 31 VLAG instances
  - **PVRST/MSTP with one VLAG instance belonging to multiple VLANs/STGs:** Maximum of 20 VLAG instances

Note: VLAG is not supported in RSTP mode.

Each type of aggregation can contain up to 24 member ports, depending on the port type and availability.

VLAGs versus Port LAGs

Though similar to regular port LAGs in many regards, VLAGs differ from regular port LAGs in a number of important ways:

- A VLAG can consist of multiple ports on two VLAG peers, which are connected to one logical client device such as a server, switch or another VLAG device.
- The participating ports on the client device are configured as a regular port LAG.
- The VLAG peers must be the same model and run the same software version.
- VLAG peers require a dedicated inter-switch link (ISL) for synchronization. The ports used to create the ISL must have the following properties:
  - ISL ports must have VLAN tagging turned on.
  - ISL ports must be configured for all VLAG VLANs.
  - ISL ports must be placed into a regular port LAG (dynamic or static).
  - A minimum of two ports on each switch are recommended for ISL use.
- Dynamic routing protocols, such as OSPF, cannot terminate on VLAGs.
- Routing over VLAGs is not supported. However, IP forwarding between subnets served by VLAGs can be accomplished using VRRP.
- VLAGs are configured using additional commands.
- It is recommended that end-devices connected to VLAG switches use NICs with dual-homing. This increases traffic efficiency, reduces ISL load and provides faster link failover.

Configuring VLAGs

When configuring VLAG or making changes to your VLAG configuration, consider the following VLAG behavior:

- When adding a static Mrouter on VLAG links, ensure that you also add it on the ISL link to avoid VLAG link failure. If the VLAG link fails, traffic cannot be recovered through the ISL. Also, make sure you add the same static entry on the peer VLAG switch for VLAG ports.
- When you enable VLAG on the switch, if a MSTP region mismatch is detected with the VLAG peer, the ISL will shut down. In such a scenario, correct the region on the VLAG peer and manually enable the ISL.
- If you have enabled VLAG on the switch, and you need to change the STP mode, ensure that you first disable VLAG and then change the STP mode.
- When VLAG is enabled, you may see two root ports on the secondary VLAG switch. One of these will be the actual root port for the secondary VLAG switch and the other will be a root port synced with the primary VLAG switch.
- The LACP key used must be unique for each VLAG in the entire topology.
- The STG to VLAN mapping on both VLAG peers must be identical.

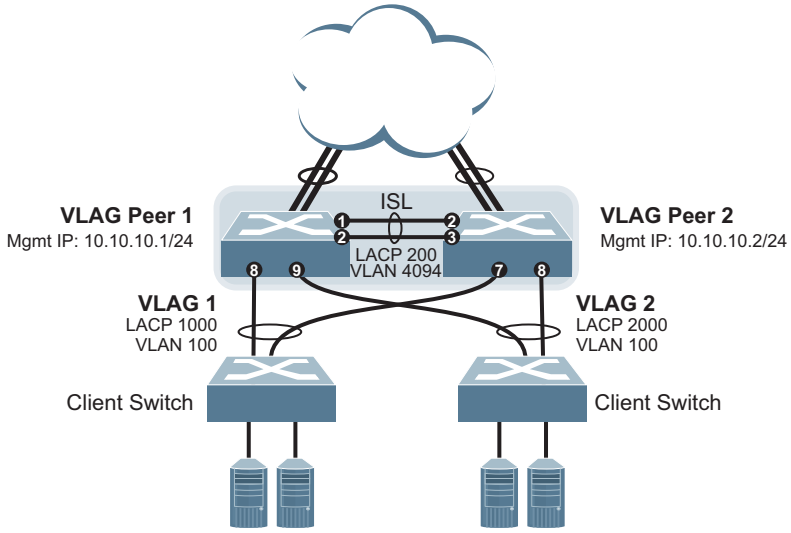
The following parameters must be identically configured on the VLAG ports of both the VLAG peers:

- VLANs
- Native VLAN tagging
- Native VLAN/PVID
- STP mode
- BPDU Guard setting
- STP port setting
- MAC aging timers
- Static MAC entries
- ACL configuration parameters
- QoS configuration parameters

Basic VLAG Configuration

Figure 19 shows an example configuration where two VLAG peers are used for aggregating traffic from downstream devices.

Figure 19. Basic VLAGs



In this example, each client switch is connected to both VLAG peers. On each client switch, the ports connecting to the VLAG peers are configured as a dynamic LACP port LAG. The VLAG peer switches share a dedicated ISL for synchronizing VLAG information. On the individual VLAG peers, each port leading to a specific client switch (and part of the client switch's port LAG) is configured as a VLAG.

In the following example configuration, only the configuration for VLAG 1 on VLAG Peer 1 is shown. VLAG Peer 2 and all other VLAGs are configured in a similar fashion.

Configure the ISL

The ISL connecting the VLAG peers is shared by all their VLAGs. The ISL needs to be configured only once on each VLAG peer.

1. Configure STP if required. Use PVRST or MSTP mode only:

```
CN 4093(config)# spanning-tree mode pvrst
```

2. Configure the ISL ports and place them into a LAG (dynamic or static):

```
CN 4093(config)# interface port 1-2
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 200
CN 4093(config-if)# exit
CN 4093(config)# vlag isl adminkey 200
```

Notes:

- In this case, a dynamic LAG is shown. A static LAG (portchannel) could be configured instead.
- ISL ports and VLAG ports must be members of the same VLANs.

3. Configure VLAG Tier ID. This is used to identify the VLAG switch in a multi-tier environment.

```
CN 4093(config)# vlag tier-id 10
```

4. Configure the ISL for the VLAG peer.

   Make sure you configure the VLAG peer (VLAG Peer 2) using the same ISL aggregation type (dynamic or static), the same VLAN and the same STP mode and tier ID used on VLAG Peer 1.

Configure the VLAG

1. Configure the VLAN for VLAG 1 ports. Once the VLAN s ready, the ISL ports are automatically added to it.

```
CN 4093(config)# vlan 100
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 8
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit
```

   Note: In MSTP mode, VLANs are automatically mapped to CIST.

2. Place the VLAG 1 port(s) in a port LAG:

```
CN 4093(config)# interface port 8
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 1000
CN 4093(config-if)# exit
```

3. Assign the LAG to the VLAG:

```
CN 4093(config)# vlag adminkey 1000 enable
```

4. Continue by configuring all required VLAGs on VLAG Peer 1 and then repeat the configuration for VLAG Peer 2.

   For each corresponding VLAG on the peer, the port LAG type (dynamic or static), the port's VLAN, and STP mode and ID must be the same as on VLAG Peer 1.

5. Enable VLAG globally.

```
CN 4093(config)# vlag enable
```

6. Verify the completed configuration:

```
CN 4093(config)# show vlag information
```

VLAG Configuration - VLANs Mapped to MSTI

Follow the steps in this section to configure VLAG in environments where the STP mode is MSTP and no previous VLAG was configured.

Configure the ISL

The ISL connecting the VLAG peers is shared by all their VLAGs. The ISL needs to be configured only once on each VLAG peer. Ensure you have the same region name, revision and VLAN-to-STG mapping on both VLAG switches.

1. Configure STP:

```
CN 4093(config)# spanning-tree mode mst
```

2. Configure the ISL ports and place them into a portchannel (dynamic or static):

```
CN 4093(config)# interface port 1-2
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 200
CN 4093(config-if)# exit
CN 4093(config)# vlag isl adminkey 200
```

Note:

a. In this case, a dynamic LAG is shown. A static LAG (portchannel) could be configured instead.

b. ISL ports and VLAG ports must be members of the same VLANs.

3. Configure VLAG Tier ID. This is used to identify the VLAG switch in a multi-tier environment.

```
CN 4093(config)# vlag tier-id 10
```

4. Configure the ISL for the VLAG peer.

   Make sure you configure the VLAG peer (VLAG Peer 2) using the same ISL aggregation type (dynamic or static), the same VLAN for vLAG ports and vLAG ISL ports, and the same STP mode and tier ID used on VLAG Peer 1.

Configure the VLAG

1. Configure the VLAN for VLAG 1 ports. Once the VLAN s ready, the ISL ports are automatically added to it.

```
CN 4093(config)# vlan 100
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 8
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit
```

2. Map the VLAN to an MSTI.

```
CN 4093(config)# spanning-tree mst configuration
CN 4093(config-mst)# instance 1 vlan 100
```

3. Place the VLAG 1 port(s) in a LAG (static or dynamic) and assign it to the VLAG:

```
CN 4093(config)# interface port 8
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 1000
CN 4093(config-if)# exit
CN 4093(config)# vlag adminkey 1000 enable
```

4. Enable VLAG:

```
CN 4093(config)# vlag enable
```

5. Continue by configuring all required VLAGs on VLAG Peer 1, and then follow the steps for configuring VLAG Peer 2.

   For each corresponding VLAG on the peer, the port LAG type (dynamic or static), the port's VLAN and STP mode and ID must be the same as on VLAG Peer 1.

6. Verify the completed configuration:

```
CN 4093# show vlag information
```

Configuring Health Check

We strongly recommend that you configure the CN4093 to check the health status of its VLAG peer. Although the operational status of the VLAG peer is generally determined via the ISL connection, configuring a network health check provides an alternate means to check peer status in case the ISL links fail. Use an independent link between the VLAG switches to configure health check.

Note: Configuring health check on an ISL VLAN interface or on a VLAG data port may impact the accuracy of the health check status.

1. Configure a management interface for the switch.

   Note: If the switch does not have a dedicated management interface, configure a VLAN for the health check interface. The health check interface can be configured with an IPv4 or IPv6 address:

```
CN 4093(config)# interface ip 127
CN 4093(config-ip-if)# ip address 10.10.10.1 255.255.255.0
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
```

   Note: Configure a similar interface on VLAG Peer 2. For example, use IP address 10.10.10.2.

2. Specify the IPv4 or IPv6 address of the VLAG Peer:

```
CN 4093(config)# vlag hlthchk peer-ip 10.10.10.2
```

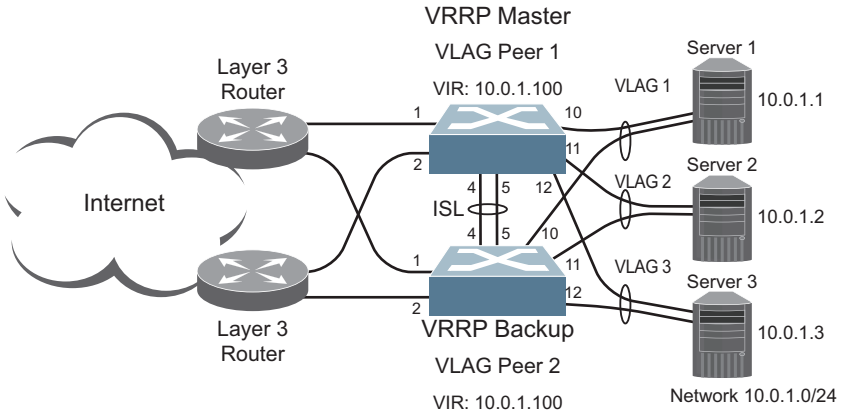
   Note: For VLAG Peer 2, the management interface would be configured as 10.10.10.2, and the health check would be configured for 10.10.10.1, pointing back to VLAG Peer 1.

   VLAG supports either IPv4 or IPv6 health check addresses at one time. Configuring an IPv4 health check address, will set any IPv6 health check address in the VLAG to 0, and vice-versa.

VLAGs with VRRP

VRRP (see "Virtual Router Redundancy Protocol" on page 527) can be used in conjunction with VLAGs and LACP-capable devices to provide seamless redundancy.

Figure 20. Active-Active Configuration using VRRP and VLAGs



Note: In a multi-layer environment, configure VRRP separately for each layer. We recommend that you configure VRRP only on the tier with uplinks. See "Configuring VLAGs in Multiple Layers" on page 216.

Configure VLAG Peer 1

1. Configure VLAG tier ID and enable VLAG globally.

```
CN 4093(config)# vlag tier-id 10
CN 4093(config)# vlag enable
```

2. Configure appropriate routing.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area 1 area-id 0.0.0.1
CN 4093(config-router-ospf)# enable
CN 4093(config-router-ospf)# exit
```

Although OSPF is used in this example, static routing could also be deployed. For more information, see "OSPF" on page 471 or "Basic IP Routing" on page 397.

3. Configure a server-facing interface.

```
CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ip address 10.0.1.10 255.255.255.0
CN 4093(config-ip-if)# vlan 100
CN 4093(config-ip-if)# exit
```

4. Turn on VRRP and configure the Virtual Interface Router.

```
CN 4093(config)# router vrrp
CN 4093(config-vrrp)# enable
CN 4093(config-vrrp)# virtual-router 1 virtual-router-id 1
CN 4093(config-vrrp)# virtual-router 1 interface 3
CN 4093(config-vrrp)# virtual-router 1 address 10.0.1.100
CN 4093(config-vrrp)# virtual-router 1 enable
```

5. Set the priority of Virtual Router 1 to 101, so that it becomes the Master.

```
CN 4093(config-vrrp)# virtual-router 1 priority 101
CN 4093(config-vrrp)# exit
```

6. Configure the ISL ports and place them into a port LAG:

```
CN 4093(config)# interface port 4-5
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 2000
CN 4093(config-if)# exit
```

Note: In this case, a dynamic LAG is shown. A static LAG could be configured instead.

7. Configure the upstream ports.

```
CN 4093(config)# interface port 1
CN 4093(config-if)# switchport access vlan 10
CN 4093(config-if)# exit
CN 4093(config)# interface port 2
CN 4093(config-if)# switchport access vlan 20
CN 4093(config-if)# exit
```

8. Configure the server ports.

```
CN 4093(config)# interface port 10
CN 4093(config-if)# switchport access vlan 100
CN 4093(config-if)# exit
CN 4093(config)# interface port 11
CN 4093(config-if)# switchport access vlan 100
CN 4093(config-if)# exit
CN 4093(config)# interface port 12
CN 4093(config-if)# switchport access vlan 100
CN 4093(config-if)# exit
```

9. Configure all VLANs including VLANs for the VLAGs.

```
CN 4093(config)# vlan 10
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 1
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit

CN 4093(config)# vlan 20
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 2
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit

CN 4093(config)# vlan 100
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 4-5, 10-12
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit
```

10. Configure Internet-facing interfaces.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 172.1.1.10 255.255.255.0
CN 4093(config-ip-if)# vlan 10
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# ip ospf area 1
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 172.1.3.10 255.255.255.0
CN 4093(config-ip-if)# vlan 20
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# ip ospf area 1
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

11. Place the VLAG port(s) in their port LAGs.

```
CN 4093(config)# interface port 10
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 1000
CN 4093(config-if)# exit
CN 4093(config)# interface port 11
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 1100
CN 4093(config-if)# exit
CN 4093(config)# interface port 12
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 1200
CN 4093(config-if)# exit
```

12. Assign the LAGs to the VLAGs:

```
CN 4093(config)# vlag adminkey 1000 enable
CN 4093(config)# vlag adminkey 1100 enable
CN 4093(config)# vlag adminkey 1200 enable
```

13. Verify the completed configuration:

```
CN 4093(config)# show vlag
```

Configure VLAG Peer 2

The VLAG peer (VLAG Peer 2) must be configured using the same ISL aggregation type (dynamic or static), the same VLAN and the same STP mode and Tier ID used on VLAG Switch 1.

For each corresponding VLAG on the peer, the port LAG type (dynamic or static), VLAN and STP mode and ID must be the same as on VLAG Switch 1.

1. Configure VLAG tier ID and enable VLAG globally.

```
CN 4093(config)# vlag tier-id 10
CN 4093(config)# vlag enable
```

2. Configure appropriate routing.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area 1 area-id 0.0.0.1
CN 4093(config-router-ospf)# enable
CN 4093(config-router-ospf)# exit
```

Although OSPF is used in this example, static routing could also be deployed.

3. Configure a server-facing interface.

```
CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ip address 10.0.1.11 255.255.255.0
CN 4093(config-ip-if)# vlan 100
CN 4093(config-ip-if)# exit
```

4. Turn on VRRP and configure the Virtual Interface Router.

```
CN 4093(config)# router vrrp
CN 4093(config-vrrp)# enable
CN 4093(config-vrrp)# virtual-router 1 virtual-router-id 1
CN 4093(config-vrrp)# virtual-router 1 interface 3
CN 4093(config-vrrp)# virtual-router 1 address 10.0.1.100
CN 4093(config-vrrp)# virtual-router 1 enable
```

5. Configure the ISL ports and place them into a port LAG:

```
CN 4093(config)# interface port 4-5
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 2000
CN 4093(config-if)# exit
```

6. Configure the upstream ports.

```
CN 4093(config)# interface port 1
CN 4093(config-if)# switchport access vlan 30
CN 4093(config-if)# exit
CN 4093(config)# interface port 2
CN 4093(config-if)# switchport access vlan 40
CN 4093(config-if)# exit
```

7. Configure the server ports.

```
CN 4093(config)# interface port 10
CN 4093(config-if)# switchport access vlan 100
CN 4093(config-if)# exit
CN 4093(config)# interface port 11
CN 4093(config-if)# switchport access vlan 100
CN 4093(config-if)# exit
CN 4093(config)# interface port 12
CN 4093(config-if)# switchport access vlan 100
CN 4093(config-if)# exit
```

8. Configure all VLANs including VLANs for the VLAGs.

```
CN 4093(config)# vlan 30
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 1
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit

CN 4093(config)# vlan 40
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 2
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit

CN 4093(config)# vlan 100
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 4-5,10-12
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit
```

9. Configure Internet-facing interfaces.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 172.1.2.11 255.255.255.0
CN 4093(config-ip-if)# vlan 30
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# ip ospf area 1
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 172.1.4.12 255.255.255.0
CN 4093(config-ip-if)# vlan 40
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# ip ospf area 1
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

10. Place the VLAG port(s) in their port LAGs:

```
CN 4093(config)# interface port 10
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 1000
CN 4093(config-if)# exit
CN 4093(config)# interface port 11
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 1100
CN 4093(config-if)# exit
CN 4093(config)# interface port 12
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# lacp key 1200
CN 4093(config-if)# exit
```

11. Assign the LAGs to the VLAGs:

```
CN 4093(config)# vlag adminkey 1000 enable
CN 4093(config)# vlag adminkey 1100 enable
CN 4093(config)# vlag adminkey 1200 enable
```

12. Verify the completed configuration:

```
CN 4093(config)# show vlag
```

Two-tier vLAGs with VRRP

vLAG Active-Active VRRP makes the secondary vLAG switch route Layer 3 traffic, thus reducing routing latency. If it is used in a two-tier vLAG environment, there may be two VRRP master switches for one VRRP domain and their role will constantly flap. To prevent such occurrences, there are two vLAG VRRP modes:

1. vLAG VRRP Active (Full Active-Active) mode

In `active` mode, Layer 3 traffic is forwarded in all vLAG related VRRP domains. To enable vLAG VRRP active mode on a switch, use the following command:

```
CN 4093(config)# vlag vrrp active
```

Note: This is the default vLAG VRRP mode.

2. vLAG VRRP Passive (Half Active-Active) mode

In `passive` mode, Layer 3 traffic is forwarded in a vLAG related VRRP domain only if either the switch or its peer virtual router is the VRRP master. To enable vLAG VRRP passive mode on a switch, use the following command:

```
CN 4093(config)# no vlag vrrp active
```

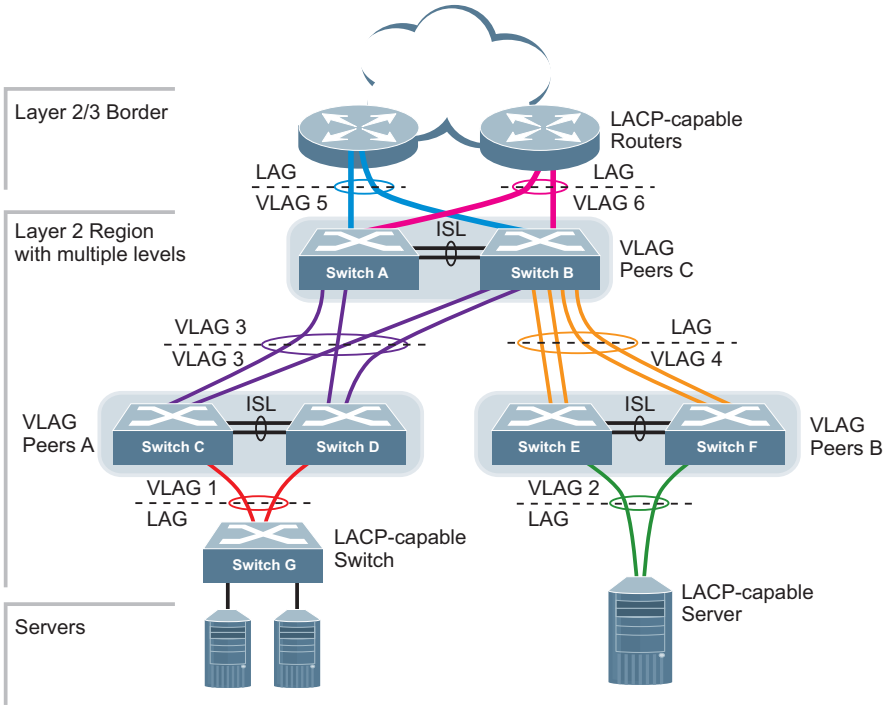
To verify the currently configured vLAG VRRP mode you can use the following command:

```
CN 4093(config)# show vlag vrrp
```

Configuring VLAGs in Multiple Layers

Figure 21 shows an example of VLAG being used in a multi-layer environment. Following are the configuration steps for the topology.

Figure 21. VLAG in Multiple Layers



Configure Layer 2/3 Border Switches

Configure ports on border switch as follows:

```
CN 4093(config)# interface port 1,2
CN 4093(config-if)# lacp key 100
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# exit
```

Repeat these steps for the second border switch.

Configure Switches in the Layer 2 Region

Consider the following:

- ISL ports on switches A and B - ports 1, 2
- Ports connecting to Layer 2/3 - ports 5, 6
- Ports on switches A and B connecting to switches C and D: ports 10, 11
- Ports on switch B connecting to switch E: ports 15, 16
- Ports on switch B connecting to switch F: ports 17, 18

1. Configure VLAG tier ID and enable VLAG globally.

```
CN 4093(config)# vlag tier-id 10
CN 4093(config)# vlag enable
```

2. Configure ISL ports on Switch A.

```
CN 4093(config)# interface port 1,2
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# lacp key 200
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# exit
CN 4093(config)# vlag isl adminkey 200
CN 4093(config-vlan)# exit
```

3. Configure port on Switch A connecting to Layer 2/3 router 1.

```
CN 4093(config)# vlan 10
 VLAN number 10 with name “VLAN 10” created
 VLAN 10 was assigned to STG 10
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 1,2,5
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit
CN 4093(config)# interface port 5
CN 4093(config-if)# lacp key 400
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# exit

CN 4093(config)# vlag adminkey 400 enable
```

Repeat these steps on Switch B for ports connecting to Layer 2/3 router 1.

4. Configure port on Switch A connecting to Layer 2/3 router 2.

```
CN 4093(config)# vlan 20
 VLAN number 20 with name “VLAN 20” created
 VLAN 20 was assigned to STG 20
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 1,2,6
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit

CN 4093(config)# interface port 6
CN 4093(config-if)# lacp key 500
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# exit

CN 4093(config)# vlag adminkey 500 enable
```

Repeat these steps on Switch B for ports connecting to Layer 2/3 router 2.

5. Configure ports on Switch A connecting to downstream VLAG switches C and D.

```
CN 4093(config)# vlan 20
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 10,11
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# lacp key 600
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# exit

CN 4093(config)# vlag adminkey 600 enable
```

   Repeat these steps on Switch B for ports connecting to downstream VLAG switch C and D.

6. Configure ports on Switch B connecting to downstream switches E and F.

```
CN 4093(config)# vlan 30
CN 4093(config-vlan)# exit
CN 4093(config)# interface port 15-18
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# lacp key 700
CN 4093(config-if)# lacp mode active
CN 4093(config-if)# exit
```

7. Configure ISL between switches C and D, and between E and F as shown in Step 1.

8. Configure the Switch G as shown in Step 2.

Chapter 12. Quality of Service

Quality of Service (QoS) features allow you to allocate network resources to mission-critical applications at the expense of applications that are less sensitive to such factors as time delays or network congestion. You can configure your network to prioritize specific types of traffic, ensuring that each type receives the appropriate QoS level.

The following topics are discussed in this section:

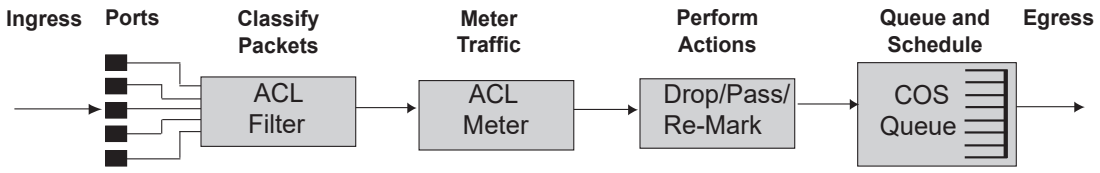
QoS Overview

QoS helps you allocate guaranteed bandwidth to critical applications, and limit bandwidth for less critical applications. Applications such as video and voice must have a certain amount of bandwidth to work correctly; using QoS, you can provide that bandwidth when necessary. Also, you can put a high priority on applications that are sensitive to timing out or those that cannot tolerate delay, assigning that traffic to a high-priority queue.

By assigning QoS levels to traffic flows on your network, you can ensure that network resources are allocated where they are needed most. QoS features allow you to prioritize network traffic, thereby providing better service for selected applications.

Figure 22 on page 219 shows the basic QoS model used by the CN4093 10 Gb Converged Scalable Switch.

Figure 22. QoS Model



The CN4093 uses the Differentiated Services (DiffServ) architecture to provide QoS functions. DiffServ is described in IETF RFC 2474 and RFC 2475.

With DiffServ, you can establish policies for directing traffic. A policy is a traffic-controlling mechanism that monitors the characteristics of the traffic (for example, its source, destination, and protocol) and performs a controlling action on the traffic when certain characteristics are matched.

The CN4093 can classify traffic by reading the DiffServ Code Point (DSCP) or IEEE 802.1p priority value, or by using filters to match specific criteria. When network traffic attributes match those specified in a traffic pattern, the policy instructs the CN4093 to perform specified actions on each packet that passes through it. The packets are assigned to different Class of Service (COS) queues and scheduled for transmission.

The basic CN4093 QoS model works as follows:

- Classify traffic:
  - Read DSCP
  - Read 802.1p Priority
  - Match ACL filter parameters
- Meter traffic:
  - Define bandwidth and burst parameters
  - Select actions to perform on in-profile and out-of-profile traffic
- Perform actions:
  - Drop packets
  - Pass packets
  - Mark DSCP or 802.1p Priority
  - Set COS queue (with or without re-marking)
- Queue and schedule traffic:
  - Place packets in one of the available COS queues
  - Schedule transmission based on the COS queue weight

Using ACL Filters

Access Control Lists (ACLs) are filters that allow you to classify and segment traffic, so you can provide different levels of service to different traffic types. Each filter defines conditions that packets must match for inclusion in a particular service class, and also the actions that are performed for matching traffic.

The CN4093 allows you to classify packets based on various parameters. For example:

- Ethernet—source MAC, destination MAC, VLAN number/mask, Ethernet type, priority
- IPv4—source IP address/mask, destination address/mask, type of service, IP protocol number
- IPv6—source IP address/prefix, destination address/prefix, next header, flow label, traffic class
- TCP/UPD—source port, destination port, TCP flag
- Packet format—Ethernet format, tagging format, IPv4, IPv6
- Egress port

For ACL details, see "Access Control Lists" on page 125.

Summary of ACL Actions

Actions determine how the traffic is treated. The CN4093 QoS actions include the following:

- Pass or Drop the packet
- Re-mark the packet with a new DiffServ Code Point (DSCP)
- Re-mark the 802.1p field
- Set the COS queue

ACL Metering and Re-Marking

You can define a profile for the aggregate traffic flowing through the CN4093 by configuring a QoS meter (if desired) and assigning ACL Groups to ports. When you add ACL Groups to a port, make sure they are ordered correctly in terms of precedence.

Actions taken by an ACL are called *In-Profile* actions. You can configure additional In-Profile and Out-of-Profile actions on a port. Data traffic can be metered, and re-marked to ensure that the traffic flow provides certain levels of service in terms of bandwidth for different types of network traffic.

Metering

QoS metering provides different levels of service to data streams through user-configurable parameters. A meter is used to measure the traffic stream against a traffic profile which you create. Thus, creating meters yields In-Profile and Out-of-Profile traffic for each ACL, as follows:

- **In-Profile**–If there is no meter configured or if the packet conforms to the meter, the packet is classified as In-Profile.
- **Out-of-Profile**–If a meter is configured and the packet does not conform to the meter (exceeds the committed rate or maximum burst rate of the meter), the packet is classified as Out-of-Profile.

Note: Metering is not supported for IPv6 ACLs. All traffic matching an IPv6 ACL is considered in-profile for re-marking purposes.

Using meters, you set a Committed Rate in Kbps (1000 bits per second in each Kbps). All traffic within this Committed Rate is In-Profile. Additionally, you can set a Maximum Burst Size that specifies an allowed data burst larger than the Committed Rate for a brief period. These parameters define the In-Profile traffic.

Meters keep the sorted packets within certain parameters. You can configure a meter on an ACL, and perform actions on metered traffic, such as packet re-marking.

Re-Marking

Re-marking allows for the treatment of packets to be reset based on new network specifications or desired levels of service. You can configure the ACL to re-mark a packet as follows:

- Change the DSCP value of a packet, used to specify the service level traffic should receive.
- Change the 802.1p priority of a packet.

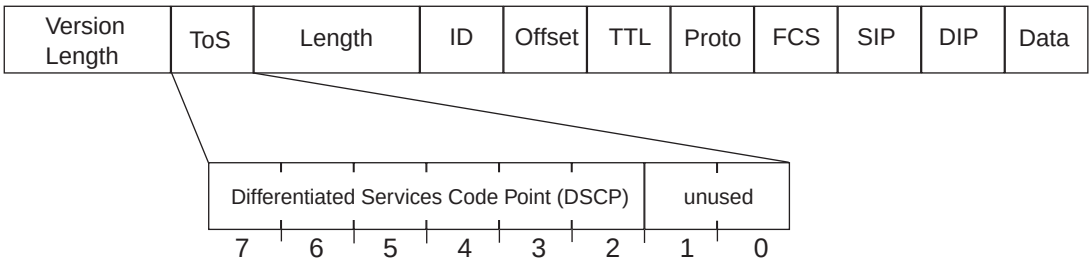
Using DSCP Values to Provide QoS

The six most significant bits in the TOS byte of the IP header are defined as DiffServ Code Points (DSCP). Packets are marked with a certain value depending on the type of treatment the packet must receive in the network device. DSCP is a measure of the Quality of Service (QoS) level of the packet.

Differentiated Services Concepts

To differentiate between traffic flows, packets can be classified by their DSCP value. The Differentiated Services (DS) field in the IP header is an octet, and the first six bits, called the DS Code Point (DSCP), can provide QoS functions. Each packet carries its own QoS state in the DSCP. There are 64 possible DSCP values (0-63).

Figure 23. Layer 3 IPv4 Packet



The CN4093 can perform the following actions to the DSCP:

- Read the DSCP value of ingress packets
- Re-mark the DSCP value to a new value
- Map the DSCP value to an 802.1p priority

Once the DSCP value is marked, the CN4093 can use it to direct traffic prioritization.

Per-Hop Behavior

The DSCP value determines the Per Hop Behavior (PHB) of each packet. The PHB is the forwarding treatment given to packets at each hop. QoS policies are built by applying a set of rules to packets, based on the DSCP value, as they hop through the network.

The CN4093 default settings are based on the following standard PHBs, as defined in the IEEE standards:

- Expedited Forwarding (EF)—This PHB has the highest egress priority and lowest drop precedence level. EF traffic is forwarded ahead of all other traffic. EF PHB is described in RFC 2598.
- Assured Forwarding (AF)—This PHB contains four service levels, each with a different drop precedence, as shown below. Routers use drop precedence to determine which packets to discard last when the network becomes congested. AF PHB is described in RFC 2597.

Drop Precedence	Class 1	Class 2	Class 3	Class 4
Low	AF11 (DSCP 10)	AF21 (DSCP 18)	AF31 (DSCP 26)	AF41 (DSCP 34)
Medium	AF12 (DSCP 12)	AF22 (DSCP 20)	AF32 (DSCP 28)	AF42 (DSCP 36)
High	AF13 (DSCP 14)	AF23 (DSCP 22)	AF33 (DSCP 30)	AF43 (DSCP 38)

- Class Selector (CS)—This PHB has eight priority classes, with CS7 representing the highest priority, and CS0 representing the lowest priority, as shown below. CS PHB is described in RFC 2474.

Priority	Class Selector	DSCP
Highest	CS7	56
	CS6	48
	CS5	40
	CS4	32
	CS3	24
	CS2	16
	CS1	8
Lowest	CS0	0

QoS Levels

Table 18 shows the default service levels provided by the CN4093, listed from highest to lowest importance:

Table 18. *Default QoS Service Levels*

Service Level	Default PHB	802.1p Priority
Critical	CS7	7
Network Control	CS6	6
Premium	EF, CS5	5
Platinum	AF41, AF42, AF43, CS4	4
Gold	AF31, AF32, AF33, CS3	3
Silver	AF21, AF22, AF23, CS2	2
Bronze	AF11, AF12, AF13, CS1	1
Standard	DF, CS0	0

DSCP Re-Marking and Mapping

The CN4093 can re-mark the DSCP value of ingress packets to a new value, and set the 802.1p priority value, based on the DSCP value. You can view the settings by using the following command:

```
CN 4093(config)# show qos dscp
Current DSCP Remarking Configuration: OFF

  DSCP    New DSCP  New 802.1p Prio
--------  --------  ---------------
    0         0          0
    1         1          0
...
   57        57          0
   58        58          0
   59        59          0
   60        60          0
   61        61          0
   62        62          0
   63        63          0
```

Use the following command to turn on DSCP re-marking globally:

```
CN 4093(config)# qos dscp re-marking
```

Then you must enable DSCP re-marking on any port that you wish to perform this function.

Note: If an ACL meter is configured for DSCP re-marking, the meter function takes precedence over QoS re-marking.

DSCP Re-Marking Configuration Example 1

The following example includes the basic steps for re-marking DSCP value and mapping DSCP value to 802.1p.

1. Turn DSCP re-marking on globally, and define the DSCP-DSCP-802.1p mapping. You can use the default mapping.

```
CN 4093(config)# qos dscp re-marking
CN 4093(config)# qos dscp dscp-mapping <DSCP value (0-63)>  <new value>
CN 4093(config)# qos dscp dot1p-mapping <DSCP value (0-63)>  <802.1p value>
```

2. Enable DSCP re-marking on a port.

```
CN 4093(config)# interface port 1
CN 4093(config-if)# qos dscp re-marking
CN 4093(config-if)# exit
```

DSCP Re-Marking Configuration Example 2

The following example assigns strict priority to VoIP traffic and a lower priority to all other traffic.

1. Create an ACL to re-mark DSCP value and COS queue for all VoIP packets.

```
CN 4093(config)# access-control list 2 tcp-udp source-port 5060 0xffff
CN 4093(config)# access-control list 2 meter committed-rate 10000000
CN 4093(config)# access-control list 2 meter enable
CN 4093(config)# access-control list 2 re-mark in-profile dscp 56
CN 4093(config)# access-control list 2 re-mark dot1p 7
CN 4093(config)# access-control list 2 action permit
```

2. Create an ACL to set a low priority to all other traffic.

```
CN 4093(config)# access-control list 3 action set-priority 1
CN 4093(config)# access-control list 3 action permit
```

3. Apply the ACLs to a port and enable DSCP marking.

```
CN 4093(config)# interface port 5
CN 4093(config-if)# access-control list 2
CN 4093(config-if)# access-control list 3 ethernet source-mac-address
00:00:00:00:00:00 00:00:00:00:00:00
CN 4093(config-if)# dscp-marking
CN 4093(config-if)# exit
```

4. Enable DSCP re-marking globally.

```
CN 4093(config)# qos dscp re-marking
```

5. Assign the DSCP re-mark value.

```
CN 4093(config)# qos dscp dscp-mapping 40 9
CN 4093(config)# qos dscp dscp-mapping 46 9
```

6. Assign strict priority to VoIP COS queue.

```
CN 4093(config)# qos transmit-queue weight-cos 7 0
```

7. Map priority value to COS queue for non-VoIP traffic.

```
CN 4093(config)# qos transmit-queue mapping 1 1
```

8. Assign weight to the non-VoIP COS queue.

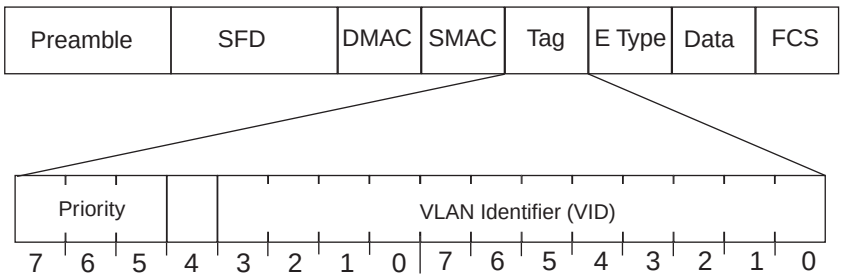
```
CN 4093(config)# qos transmit-queue weight-cos 1 2
```

Using 802.1p Priorities to Provide QoS

Enterprise NOS provides Quality of Service functions based on the priority bits in a packet's VLAN header. (The priority bits are defined by the 802.1p standard within the IEEE 802.1q VLAN header.) The 802.1p bits, if present in the packet, specify the priority that should be given to packets during forwarding. Packets with a numerically higher (non-zero) priority are given forwarding preference over packets with lower priority bit value.

The IEEE 802.1p standard uses eight levels of priority (0-7). Priority 7 is assigned to highest priority network traffic, such as OSPF or RIP routing table updates, priorities 5-6 are assigned to delay-sensitive applications such as voice and video, and lower priorities are assigned to standard applications. A value of 0 (zero) indicates a "best effort" traffic prioritization, and this is the default when traffic priority has not been configured on your network. The CN4093 can filter packets based on the 802.1p values, and it can assign or overwrite the 802.1p value in the packet.

Figure 24. Layer 2 802.1q/802.1p VLAN Tagged Packet



Ingress packets receive a priority value, as follows:

- **Tagged packets**—CN4093 reads the 802.1p priority in the VLAN tag.
- **Untagged packets**—CN4093 tags the packet and assigns an 802.1p priority, based on the port's default priority.

Egress packets are placed in a COS queue based on the priority value, and scheduled for transmission based on the scheduling weight of the COS queue.

To configure a port's default 802.1p priority value, use the following commands.

```
CN 4093(config)# interface port 1
CN 4093(config-if)# dot1p <802.1p value (0-7)>
CN 4093(config-if)# exit
```

See "Queuing and Scheduling" on page 229 for details on scheduling weights.

Queuing and Scheduling

The CN4093 can be configured to have either 2 or 8 output Class of Service (COS) queues per port, into which each packet is placed. Each packet's 802.1p priority determines its COS queue, except when an ACL action sets the COS queue of the packet.

You can configure the following attributes for COS queues:

- Map 802.1p priority value to a COS queue
- Define the scheduling weight of each COS queue

You can map 802.1p priority value to a COS queue, as follows:

```
CN 4093(config)# qos transmit-queue mapping <priority value (0-7)> <COS queue (0-7)>
```

To set the COS queue scheduling weight, use the following command.

```
CN 4093(config)# qos transmit-queue weight-cos <COSq number> <COSq weight (0-15)>
```

The scheduling weight can be set from 0 to 15. Weight values from 1 to 15 set the queue to use weighted round-robin (WRR) scheduling, which distributes larger numbers of packets to queues with the highest weight values. For distribution purposes, each packet is counted the same, regardless of the packet's size.

A scheduling weight of 0 (zero) indicates strict priority. Traffic in strict priority queue has precedence over other all queues. If more than one queue is assigned a weight of 0, the strict queue with highest queue number will be served first. Once all traffic in strict queues is delivered, any remaining bandwidth will be allocated to the WRR queues, divided according to their weight values.

Note: Use caution when assigning strict scheduling to queues. Heavy traffic in queues assigned with a weight of 0 can starve lower priority queues.

For a scheduling method that uses a weighted deficit round-robin (WDRR) algorithm, distributing packets with an awareness of packet size, see "Enhanced Transmission Selection" on page 358.

Control Plane Protection

Control plane receives packets that are required for the internal protocol state machines. This type of traffic is usually received at low rate. However, in some situations such as DOS attacks, the switch may receive this traffic at a high rate. If the control plane protocols are unable to process the high rate of traffic, the switch may become unstable.

The control plane receives packets that are channeled through protocol-specific packet queues. Multiple protocols can be channeled through a common packet queue. However, one protocol cannot be channeled through multiple packet queues. These packet queues are applicable only to the packets received by the software and does not impact the regular switching or routing traffic. Packet queue with a higher number has higher priority.

You can configure the bandwidth for each packet queue. Protocols that share a packet queue will also share the bandwidth.

The following commands configure the control plane protection (CoPP) feature:

```
Configure a queue for a protocol:
CN 4093(config)# qos protocol-packet-control packet-queue-map <0-47>
<protocol>

Set the bandwidth for the queue, in packets per second:
CN 4093(config)# qos protocol-packet-control rate-limit-packet-queue
<0-47> <1-10000>
```

Packet Drop Logging

Packet drop logging allows you to monitor network deficiencies by generating syslog messages for packet drops in the CPU queues. By default, the switch will generate such messages once every 30 minutes, specifying the type of traffic, queue data rate and queue number on which the drops occurred, such as:

```
Apr 19 11:27:35 172.31.37.200 NOTICE Protocol control discards: ARP
Broadcast packets are received at rate higher than 200pps, hence are
discarded on queue 5.
```

To enable or disable packet drop logging, use the following commands:

```
CN 4093(config)# [no] logging pdrop enable
```

You can adjust the logging interval between 0 and 30 minutes using the following command:

```
CN 4093(config)# logging pdrop interval <0-30>
```

Setting the logging interval to 0 will log packet drops immediately (with up to 1 second delay), and will ignore further drops on the same queue during the next 2 minutes.

Setting the logging interval to a greater value (1 – 30 minutes), regularly displays packet drop information at the designated time intervals. Once the packet drops stop, or if new packet drops are encountered only within 2 minutes after a syslog message, the switch does not display any more messages.

Part 4: Advanced Switching Features

Chapter 13. Stacking

This chapter describes how to implement the stacking feature in the Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch. The following concepts are covered:

Stacking Overview

A hybrid *stack* is a group of eight switches: two CN4093 10 Gb Converged Scalable Switches and six EN4093R 10Gb Scalable Switches. A *stack* can also be formed with just two CN4093 10 Gb Converged Scalable Switches.

A stack has the following properties, regardless of the number of switches included:

- The network views the stack as a single entity.
- The stack can be accessed and managed as a whole using standard switch IP interfaces configured with IPv4 addresses.
- The CLI for Individual Member switches is available via the Master switch serial console or using remote Telnet/SSH access to the Master.
- Once the stacking links have been established (see the next section), the number of ports available in a stack equals the total number of remaining ports of all the switches that are part of the stack.
- The number of available IP interfaces, VLANs, LAGs, LAG Links and other switch attributes are not aggregated among the switches in a stack. The totals for the stack as a whole are the same as for any single switch configured in stand-alone mode. A maximum of 4095 VLANs are supported in stand-alone mode, and a maximum of 1024 VLANs are supported in stacking mode.

Stacking Requirements

Before Enterprise NOS switches can form a stack, they must meet the following requirements:

- Switches in a hybrid stack must be of the model CN4093 10 Gb Converged Scalable Switch or EN4093R 10Gb Scalable Switch.
- In a hybrid stack, the EN4093R switches cannot act as Backup switches. You must use only the CN4093 10Gb Converged Scalable switches as the Master switch and Backup switch.
- In a hybrid stack, only two CN4093 10Gb Converged Scalable switches can be grouped with the EN4093R switches.
- Each switch must be installed with ENOS, version 8.4. Please see "Upgrading Software in a Stack" on page 252.
- The recommended stacking topology is a bidirectional ring (see Figure 25 on page 246). To achieve this, two 10Gb or two 40 Gb Ethernet ports on each switch must be reserved for stacking. By default, 10Gb Ethernet ports EXT1 and EXT2 are used.
- Stack with only CN4093 switches supports stack LAG links that can be configured as follows:
  - Stack of two units: Maximum of two 10Gb ports or two 40 Gb ports.
  - Omni ports cannot be used as stack LAG links.
  - You cannot combine 10Gb ports with 40Gb ports in the stack LAGs.
  - An LACP port cannot be a stack LAG member. If you need to use the port in the stack LAG, you must first set the LACP port mode to `off` (`CN 4093(config-if)# lacp mode off`).
- The cables used for connecting the switches in a stack carry low-level, inter-switch communications as well as cross-stack data traffic critical to shared switching functions. Always maintain the stability of stack links to avoid internal stack reconfiguration.

Stacking Limitations

The CN4093 with ENOS 8.4 can operate in one of two modes:

- Default mode, which is the regular stand-alone (or non-stacked) mode.
- Stacking mode, in which multiple physical switches aggregate functions as a single switching device.

When in stacking mode, the following stand-alone features are not supported:

- Border Gateway Protocol (BGP)
- Ethernet Operation, Administration and Maintenance (OAM)
- Internet Group Management Protocol version 3 (IGMPv3)
- Internet Group Management Protocol (IGMP) Querier
- Internet Group Management Protocol (IGMP) Relay
- Internet Key Exchange version 2 (IKEv2)

- IP Security (IPsec)
- Internet Protocol version 6 (IPv6)
- Loopback Interfaces
- MAC address notification
- Multicast Listener Discovery (MLD)
- Network Configuration (NETCONF) Protocol
- Open Shortest Path First (OSPF)
- Open Shortest Path First version 3 (OSPFv3)
- Port flood blocking
- Protocol-based VLANs
- Router IDs
- Route maps
- Routing Information Protocol (RIP)
- sFlow port monitoring
- Spanning Tree Protocol (STP) Root Guard and Loop Guard
- Static MAC address adding
- Static Multicast Routes
- Storm control
- Switch Partition (SPAR)
- Uni-Directional Link Detection (UDLD)
- Virtual Link Aggregation Groups (VLAG)
- Virtual Router Redundancy Protocol (VRRP)

Note: In stacking mode, switch menus and command for unsupported features may be unavailable or may have no effect on switch operation.

Stack Membership

A stack contains up to eight switches, interconnected by a stack LAG in a local ring topology (see Figure 25 on page 246). With this topology, only a single stack link failure is allowed.

An operational stack must contain one Master and one or more Members, as follows:

- **Master**

  One switch controls the operation of the stack and is called the Master. The Master provides a single point to manage the stack. A stack must have one and only one Master. The firmware image, configuration information, and run-time data are maintained by the Master and pushed to each switch in the stack as necessary.

- **Member**

  Member switches provide additional port capacity to the stack. Members receive configuration changes, run-time information, and software updates from the Master.

- **Backup**

  One member switch can be designated as a Backup to the Master. The Backup takes over control of the stack if the Master fails. Configuration information and run-time data are synchronized with the Master.

The Master Switch

An operational stack can have only one active Master at any given time. In a normal stack configuration, one switch is configured as a Master and all others are configured as Members.

When adding new switches to an existing stack, the administrator must explicitly configure each new switch for its intended role as a Master (only when replacing a previous Master) or as a Member. All stack configuration procedures in this chapter depict proper role specification.

However, there are scenarios when the two stacks (each having a Master switch) are interconnected through stacking links. When this occurs, one Master switch will automatically be chosen as the active Master for the entire stack. The selection process is designed to promote stable, predictable stack operation and minimize stack reboots and other disruptions.

Splitting and Merging One Stack

If stack links or Member switches fail, any Member which cannot access either the Master or Backup is considered *isolated* and will not process network traffic (see "No Backup" on page 243). Members which have access to a Master or Backup (or both), despite other link or Member failures, will continue to operate as part of their active stack. A Member that is isolated due to link failure resets itself. After it is up, if the link failure still exits, the Member stays in isolated state keeping all its data links disabled. Only the management and stacking links are enabled. If the Member was not configured when it went to isolated state, the Master pushes the configuration when the Member joins back the stack.

If multiple stack links or stack Member switches fail, thereby separating the Master and Backup into separate sub-stacks, the Backup automatically becomes an active Master for the partial stack in which it resides. Later, if the topology failures are corrected, the partial stacks will merge, and the two active Masters will come into contact.

In this scenario, if both the (original) Master and the Backup (acting as Master) are in operation when the merger occurs, the original Master will reassert its role as active Master for the entire stack. If any configuration elements were changed and applied on the Backup during the time it acted as Master (and forwarded to its connected Members), the Backup and its affected Members will reboot and will be reconfigured by the returning Master before resuming their regular roles.

Note: When the Backup becomes a Master, if NTP is enabled from the CMM, configuration changes are made to the Backup (acting as Master) by the CMM. Therefore, in the event of a subsequent merger, the aforementioned reboot and reconfiguration of the Backup and its affected Members will occur.

However, if the original Master switch is disrupted (powered down or in the process of rebooting) when it is reconnected with the active stack, the Backup (acting as Master) will retain its acting Master status to avoid disruption to the functioning stack. The deferring Master will temporarily assume a role as Backup.

If both the Master and Backup are rebooted, all member switches in the stack will also reboot. When the switches resume operation, they will assume their originally configured roles.

If, while the stack is still split, the Backup (acting as Master) is explicitly reconfigured to become a regular Master, then when the split stacks are finally merged, the Master with the lowest MAC address will become the new active Master for the entire stack.

Merging Independent Stacks

If switches from different stacks are linked together in a stack topology without first reconfiguring their roles as recommended, it is possible that more than one switch in the stack might be configured as a Master.

Since up to 16 units can be attached to a stack, a merge between two 8 unit stack can be performed. The user will then have to choose which units will remain in the final stack and which will be eliminated, since only 8 of them can forward networking traffic, the rest having the data links disabled.

Note: Do not merge hybrid stacks if the total number of CN4093 switches exceeds two units.

Although all switches which are configured for stacking and joined by stacking links are recognized as potential stack participants by any operational Master switches, they are not brought into operation within the stack until explicitly assigned (or "bound") to a specific Master switch.

Consider two independent stacks, Stack A and Stack B, which are merged into one stacking topology. The stacks will behave independently until the switches in Stack B are bound to Master A (or vice versa). In this example, once the Stack B switches are bound to Master A, Master A will automatically reconfigure them to operate as Stack A Members, regardless of their original status within Stack B.

However, for purposes of future Backup selection, reconfigured Masters retain their identity as configured Masters, even though they otherwise act as Members. In case the configured Master goes down and the Backup takes over as the new Master, these reconfigured Masters become the new Backup. When the original configured Master of the stack boots up again, it acts as a Member. This is one way to have multiple backups in a stack.

Backup Switch Selection

An operational stack can have one optional Backup at any given time. Only the Backup specified in the active Master's configuration is eligible to take over current stack control when the Master is rebooted or fails. The Master automatically synchronizes configuration settings with the specified Backup to facilitate the transfer of control functions.

The Backup retains its status until one of the following occurs:

- The Backup switch is deleted using the following command from the Master:

```
CN 4093(config)# no stack backup
```

- The Backup switch is changed using the following command from the Master:

```
CN 4093(config)# stack backup <csnum 1-8>
```

  Note: This will replace the current Backup switch with the configured switch specified through its csnum. Even if the new Backup switch is not attached to the stack when issuing the command, the current Backup switch will loose its role. When the configured switch is attached, the command will take effect and the switch will become the Backup.

- A new Master assumes operation as active Master in the stack and uses its own configured Backup settings.
- The active Master is rebooted with the boot configuration set to factory defaults (clearing the Backup setting).

Master Failover

When the Master switch is present, it controls the operation of the stack and pushes configuration information to the other switches in the stack. If the active Master fails, then the designated Backup (if one is defined in the Master's configuration) becomes the new acting Master and the stack continues to operate normally.

Master Recovery

If the prior Master recovers in a functioning stack where the Backup has assumed stack control, the prior Master does not reassert itself as the stack Master. Instead, the prior Master will assume a role as a secondary Backup to avoid further stack disruption.

Upon stack reboot, the Master and Backup will resume their regular roles.

No Backup

If a Backup is not configured on the active Master, or the specified Backup is not operating, then if the active Master fails, the stack will reboot without an active Master.

When a group of stacked switches are rebooted without an active Master present, the switches are considered to be *isolated*. All isolated switches in the stack are placed in a `INIT` state until a Master appears. During this `INIT` period, all the network ports of these Member switches are placed into operator-disabled state. Without the Master, a stack cannot respond correctly to networking events.

Stack Member Identification

Each switch in the stack has two numeric identifiers, as follows:

- **Attached Switch Number** (`asnum`)

  An `asnum` is automatically assigned by the Master switch, based on each Member switch's physical connection in relation to the Master. The `asnum` is mainly used as an internal ID by the Master switch and is not user-configurable.

- **Configured Switch Number** (`csnum`):

  The `csnum` is the logical switch ID assigned by the stack administrator. The `csnum` is used in most stacking-related configuration commands and switch information output. It is also used as a port prefix to distinguish the relationship between the ports on different switches in the stack.

It is recommended that `asnum` 1 and `csnum` 1 be used for identifying the Master switch. By default, `csnum` 1 is assigned to the Master. If `csnum` 1 is not available, the lowest available `csnum` is assigned to the Master.

Configuring a Stack

When stacking mode is enabled on the switch, the configuration is reset to factory default and the port numbering changes.

When a switch mode is changed from stand-alone to stack or from stack to stand-alone, the active and backup configuration will be erased. We recommended that you save the configuration to an external device before changing the switch mode.

Configuration Overview

This section provides procedures for creating a stack of switches. The high-level procedure is as follows:

- Configure the stack settings to be available after the next reboot:
  - Choose one Master switch for the entire stack.
  - Set all stack switches to stacking mode.
  - Configure the same stacking VLAN for all switches in the stack.
  - Configure the desired stacking interlinks.
- Reboot the stack switches.
- Configure the stack after the reboot:
  - Bind Member switches to the Master.
  - Assign a Backup switch.

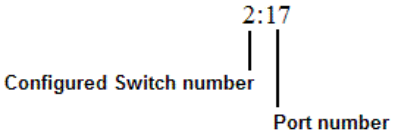
These tasks are covered in detail in the following sections.

Best Configuration Practices

The following are guidelines for building an effective switch stack:

- Always connect the stack switches in a complete ring topology (see Figure 25 on page 246).
- Avoid disrupting the stack connections unnecessarily while the stack is in operation.
- For enhanced redundancy when creating port LAGs, include ports from different stack members in the LAGs.
- Avoid altering the stack `asnum` and `csnum` definitions unnecessarily while the stack is in operation.
- When in stacking mode, the highest QoS priority queue is reserved for internal stacking requirements. Therefore, only seven priority queues will be available for regular QoS use.
- Configure only as many QoS levels as necessary. This allows the best use of packet buffers.

- Before configuring the stack:
  - Identify the VLAN to be used as the stacking VLAN.
  - Save the current configuration to an external device. The port numbering will change once stacking is enabled. Use the saved configuration to reassign ports/interfaces as per the new port numbering scheme. Once a stack is configured, port numbers are displayed throughout the BBI using the `csnum` to identify the switch, followed by the switch port number. For example:



Stacking VLANs

VLAN 4090 is the default VLAN reserved for internal traffic on stacking ports.You can change the VLAN, if required.

Note: Do not use VLAN 4090 (or the configured VLAN) for any purpose other than internal stacking traffic.

Configuring Each Switch in a Stack

To configure each switch for stacking, connect to the internal management IP interface for each switch (assigned by the management system) and use the ISCLI to perform the following steps.

Note: IPv6 is not supported in stacking mode. IP interfaces must use IPv4 addressing for proper stack configuration.

1. On each switch, enable stacking:

```
CN 4093(config)# boot stack enable
```

2. On each switch, set the stacking membership mode.

   By default, each switch is set to Member mode. However, one switch must be set to Master mode. Use the following command on only the designated Master switch:

```
CN 4093(config)# boot stack mode master
```

   Note: If any Member switches are incorrectly set to Master mode, use the `mode member` option to set them back to Member mode.

3. On each switch, configure the stacking VLAN (or use the default setting).

   Although any VLAN (except VLAN 1) may be defined for stack traffic, it is highly recommended that the default, VLAN 4090 as shown in the following example, be reserved for stacking.

```
CN 4093(config)# boot stack vlan 4090
```

4. On each switch, designate the stacking links.

If using the 2 x 40Gb ports as stacking links, first convert the 40Gb ports from their default 4x10Gb mode of operation to 40Gb mode. See: "Configuring QSFP+ Ports" on page 164.

Use the following command to specify the links to be used in the stacking LAG:

```
CN 4093(config)# boot stack higig-trunk <list of port names or aliases>
```

Note: Ports configured as Server ports for use with VMready cannot be designated as stacking links.

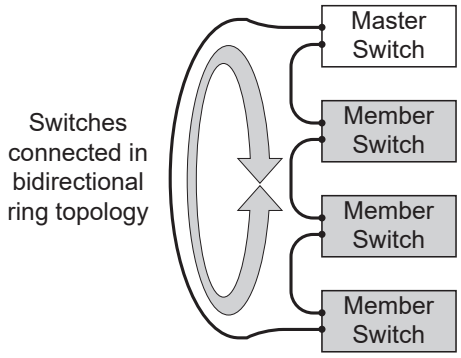
5. On each switch, perform a reboot:

```
CN 4093(config)# reload
```

6. Physically connect the stack LAGs.

   To create the recommended topology, attach the two designated stacking links in a bidirectional ring. As shown in Figure 25, connect each switch in turn to the next, starting with the Master switch. To complete the ring, connect the last Member switch back to the Master.

   Figure 25. Example of Stacking Connections



   Note: The stacking feature is designed such that the stacking links in a ring topology do not result in broadcast loops. The stacking ring is thus valid (no stacking links are blocked), even when Spanning Tree protocol is enabled.

   Once the stack LAGs are connected, the switches will perform low-level stacking configuration.

   Note: Although stack link failover/failback is accomplished on a sub-second basis, to maintain the best stacking operation and avoid traffic disruption, it is recommended not to disrupt stack links after the stack is formed.

Configuring a Management IP Interface

Each switch in a stack can be configured with the external management IP interface (127). The switch's MAC address must be associated with the management IP interface. This interface can be used for connecting to and managing the switch externally. Follow the steps below:

```
CN 4093(config)# interface ip 127
CN 4093(config-ip-if)# mac <switch MAC address> ip address <IPv4 address> <subnet mask> enable
CN 4093(config-ip-if)# exit
CN 4093(config)# ip gateway 3 mac <switch MAC address> address <gateway IPv4 address> enable
```

To provide continuous Management IP reachability in the event of a Master node failover, an additional floating Management IP address can be set up on the management interface. The floating Management IP address will be used by the backup switch when taking over management from the failed master node. To configure the floating Management IP address, use the following command:

```
CN 4093(config-if)# floating ip address <IPv4 address> <subnet mask>
```

Note: The Management IP and floating Management IP addresses on the master switch, as well as the Management IP address on the backup switch, must be in the same subnet.

Note: In case of a stack split, the floating IP cannot be used anymore due to duplicate IP address issue.

Additional Master Configuration

Once the stack links are connected, access the internal management IP interface of the Master switch (assigned by the management system) and complete the configuration.

Viewing Stack Connections

To view information about the switches in a stack, execute the following command:

```
CN 4093(config)# show stack switch

Stack name: STK
Local switch:
csnum              - 2
MAC                - 74:99:75:21:8c:00
UUID               - 98c587636548429aba5010f8c62d4e27
Bay Number         - 3
Switch Type        - 14 (CN4093)
Chassis Type       - 6 (Flex Enterprise)
Switch Mode (cfg)  - Member (backup)
Priority           - 245
Stack MAC          - 74:99:75:21:8d:1f

Master switch:
csnum              - 2
MAC                - 74:99:75:21:8c:00
UUID               - 98c587636548429aba5010f8c62d4e27
Bay Number         - 3

Backup switch:
csnum              - 1
MAC                - 74:99:75:21:8d:00
UUID               - 534c8ca1605846299148305adc9a1f6d
Bay Number         - 4

Configured Switches:
------------------------------------------------------------------------
csnum           UUID                      Bay      MAC             asnum
------------------------------------------------------------------------
C1     534c8ca1605846299148305adc9a1f6d   4   74:99:75:21:8d:00    A5
C2     98c587636548429aba5010f8c62d4e27   3   74:99:75:21:8c:00    A1
C3     534c8ca1605846299148305adc9a1f6d   1   00:00:00:00:00:00
C4     25b884f3c75341e7a0a6417d8602180b   4   08:17:f4:84:34:00    A2
C5     98c587636548429aba5010f8c62d4e27   4   34:40:b5:73:8a:00    A3
C6     534c8ca1605846299148305adc9a1f6d   3   74:99:75:1c:68:00    A4
C7     25b884f3c75341e7a0a6417d8602180b   3   00:00:00:00:00:00

Attached Switches in Stack:
------------------------------------------------------------------------------
asnum           UUID                   Bay     MAC           csnum State    Type
------------------------------------------------------------------------------
A1    98c587636548429aba5010f8c62d4e27  3  74:99:75:21:8c:00  C2   IN_STACK CN4093
A2    25b884f3c75341e7a0a6417d8602180b  4  08:17:f4:84:34:00  C4   IN_STACK CN4093
A3    98c587636548429aba5010f8c62d4e27  4  34:40:b5:73:8a:00  C5   IN_STACK CN4093
A4    534c8ca1605846299148305adc9a1f6d  3  74:99:75:1c:68:00  C6   IN_STACK CN4093
A5    534c8ca1605846299148305adc9a1f6d  4  74:99:75:21:8d:00  C1   IN_STACK CN4093
```

Binding Members to the Stack

You can bind Member switches to a stack `csnum` using either their `asnum` or chassis UUID and bay number:

```
CN 4093(config)# stack switch-number <csnum> universal-unic-id <chassis UUID>
CN 4093(config)# stack switch-number <csnum> bay <bay number (1-4)>
  -or-
CN 4093(config)# stack switch-number <csnum> bind <asnum (1-16)>
```

To remove a Member switch, execute the following command:

```
CN 4093(config)# no stack switch-number <csnum>
```

To bind all units of a stack, use the command:

```
CN 4093(config)# stack bind
```

The **`stack bind`** command automatically assigns switch numbers to all attached switches in the stack that do not yet have a number assigned.

Assigning a Stack Backup Switch

To define a Member switch as a Backup (optional) which will assume the Master role if the Master switch fails, execute the following command:

```
CN 4093(config)# stack backup <csnum>
  -or-
CN 4093(config)# stack bind
```

Managing a Stack

The stack is managed primarily through the Master switch. The Master switch then pushes configuration changes and run-time information to the Member switches.

Use Telnet or the Browser-Based Interface (BBI) to access the Master, as follows:

- Use the management IP address assigned to the Master by the management system.
- On any switch in the stack, connect to any port that is not part of an active LAG and is a member of a VLAN. To access the stack, use the IP address of any IP interface that is member of the VLAN.

Connecting to Stack Switches via the Master

From the Master switch, you can connect to any other switch in the stack directly from the ISCLI using the following command:

```
CN 4093# connect <asnum (1-16)>
```

Rebooting Stacked Switches via the Master

Rebooting Stacked Switches using the ISCLI

The administrator can reboot individual switches in the stack, or the entire stack using the following commands:

```
CN 4093(config)# reload                          (Reboot all switches in the stack)
CN 4093(config)# reload master                   (Reboot only the stack Master)
CN 4093(config)# reload switch <csnum list>      (Reboot only the listed switches)
```

Note: If no backup switch is present in the stack, the **`reload master`** command will reboot all switches.

Rebooting Stacked Switches using the BBI

The **Configure > System > Config/Image Control** window allows the administrator to perform a reboot of individual switches in the stack, or the entire stack. The following table describes the stacking Reboot buttons.

Table 19. *Stacking Boot Management buttons*

Field	Description
Reboot Stack	Performs a software reboot/reset of all switches in the stack. The software image specified in the Image To Boot drop-down list becomes the active image.
Reboot Master	Performs a software reboot/reset of the Master switch. The software image specified in the Image To Boot drop-down list becomes the active image.
Reboot Switches	Performs a reboot/reset on selected switches in the stack. Select one or more switches in the drop-down list, and click Reboot Switches. The software image specified in the Image To Boot drop-down list becomes the active image.

The **Update Image/Cfg** section of the window applies to the Master. When a new software image or configuration file is loaded, the file first loads onto the Master, and the Master pushes the file to all other switches in the stack, placing it in the same software or configuration bank as that on the Master. For example, if the new image is loaded into image 1 on the Master switch, the Master will push the same firmware to image 1 on each Member switch.

Upgrading Software in a Stack

New Hybrid Stack

Use the following procedure to install software on switches that will be used to form a hybrid stack (two CN4093 and up to six EN4093R switches):

1. Install ENOS version 8.4 on each switch and reload the switch.
2. Configure the switches to form a stack. See "Configuring a Stack" on page 244.
3. Reload the switches to establish the stack.

Converting a EN4093R Stack to a Hybrid Stack

Use the following procedure to install software on a stack of EN4093R switches that will be combined with CN4093 switches to form a hybrid stack (up to two CN4093 and up to six EN4093R switches):

1. Install ENOS version 8.4 on the Master EN4093R switch.
2. Install ENOS version 8.4 on each CN4093 switch.
3. Reload the switches.
4. Configure stacking on the CN4093 switch(es). The CN4093 must be configured as the Master of the hybrid stack. Reload the switch(es) to establish the stack.

New Stack

Use the following procedure to install software on two CN4093 switches that will be used to form a stack:

1. Install ENOS version 8.4 on each CN4093 switch and reload the switch.
2. Configure the switches to form a stack. See "Configuring a Stack" on page 244.
3. Reload the switches to establish the stack.

Replacing or Removing Stacked Switches

Stack switches may be replaced or removed while the stack is in operation. However, the following conditions must be met to avoid unnecessary disruption:

- If removing an active Master switch, make sure that a valid Backup exists in the stack.
- It is best to replace only one switch at a time.
- If replacing or removing multiple switches in a ring topology, when one switch has been properly disconnected (see the procedures that follow), any adjacent switch can also be removed.
- Removing any two, non-adjacent switches in a ring topology will divide the ring and disrupt the stack.

Use the following procedures to replace a stack switch.

Removing a Switch from the Stack

1. Make sure the stack is configured in a ring topology.

   Note: When an open-ended daisy-chain topology is in effect (either by design or as the result of any failure of one of the stacking links in a ring topology), removing a stack switch from the interior of the chain can divide the chain and cause serious disruption to the stack operation.

2. If removing a Master switch, make sure that a Backup switch exists in the stack, then turn off the Master switch.

   This will force the Backup switch to assume Master operations for the stack.

3. Remove the stack link cables from the old switch only.
4. Disconnect all network cables from the old switch only.
5. Remove the old switch.

Installing the New Switch or Healing the Topology

If using a ring topology, but not installing a new switch for the one removed, close the ring by connecting the open stack links together, essentially bypassing the removed switch.

Otherwise, if replacing the removed switch with a new unit, use the following procedure:

1. Make sure the new switch meets the stacking requirements on page 237.
2. Place the new switch in its determined place according to the *CN4093 10 Gb Converged Scalable Switch Installation Guide*.
3. Connect to the ISCLI of the new switch (not the stack interface)
4. Enable stacking:

```
CN 4093(config)# boot stack enable
```

5. Set the stacking mode.

   By default, each switch is set to Member mode. However, if the incoming switch has been used in another stacking configuration, it may be necessary to ensure the proper mode is set.

   - If replacing a Member or Backup switch:

     ```
     CN 4093(config)# boot stack mode member
     ```

   - If replacing a Master switch:

     ```
     CN 4093(config)# boot stack mode master
     ```

6. Configure the stacking VLAN on the new switch, or use the default setting.

   Although any VLAN may be defined for stack traffic, it is highly recommended that the default, VLAN 4090, be reserved for stacking, as shown in the following command.

   ```
   CN 4093(config)# boot stack vlan 4090
   ```

7. Designate the stacking links.

   Use the following command to specify the links to be used in the stacking LAG:

   ```
   CN 4093(config)# boot stack higig-trunk <list of external port s>
   ```

8. Attach the required stack link cables to the designated stack links on the new switch.
9. Attach the desired network cables to the new switch.
10. Reboot the new switch:

    ```
    CN 4093(config)# reload
    ```

    When the new switch boots, it will join the existing stack. Wait for this process to complete.

Binding the New Switch to the Stack

1. Log in to the stack interface.

   Note: If replacing the Master switch, be sure to log in to the stack interface (hosted temporarily on the Backup switch) rather than logging in directly to the newly installed Master.

2. From the stack interface, assign the `csnum` for the new switch.

   You can bind Member switches to a stack `csnum` using either the new switch's `asnum` or MAC address:

```
CN 4093(config)# stack switch-number <csnum> universal-unic-id <uuid> bay
<Slot ID>
   -or-
CN 4093(config)# stack switch-number <csnum> bind <asnum>
   -or-
CN 4093(config)# stack bind
```

   Note: If replacing the Master switch, the Master will not assume control from the Backup unless the Backup is rebooted or fails.

Performing a Rolling Reload or Upgrade

You can perform a sequential reload or upgrade, otherwise known as a staggered or *rolling* reload or upgrade, to avoid the need for an overall outage. With a rolling reload or upgrade, some of the hardware stays up at all times.

This approach differs from the traditional image upgrade that requires manual image downloads and install to individual switches, which then requires the entire logical switch reboot.

After the firmware is copied to all members of the stack, the rolling reload or upgrade process automatically reboots all switches sequentially in the following order:

- Backup switch
- Master switch
- Configured stack members, from lowest to highest `csnum`
- Attached but not configured stack members, from lowest to highest `asnum`

During the rolling firmware reload or upgrade process, there will be continuous connectivity to the upstream network. From the point of view of the stack, it is as though a series of switch and uplink failures are occurring. When the design is cabled and configured properly, the environment redirects traffic.

For detailed instructions on upgrading and rebooting, see Chapter 3, "Switch Software Management".

Starting a Rolling Reload

To start a rolling reload, use the command:

```
CN 4093(config)# reload staggered [delay <delay>]
```

where *delay* is an integer from 2 to 20 representing the time delay in minutes between switch reboots. The default value is one minute between switch reboots.

Starting a Rolling Upgrade

To start a rolling upgrade, use the command:

```
CN 4093(config)# copy {tftp|ftp|sftp} {image1|image2} {address <IP address>}
{filename <image filename>} staggered-upgrade [delay <2-20 minutes>]
```

where:

- `tftp/ftp/sftp` is the protocol for copying
- `image1/image2` is the image to which the firmware is being copied
- `address` is the IP address from which the firmware is being copied
- `filename` is the name of the firmware file that is being copied
- `delay` is the delay between each reload, in minutes

To upgrade both the boot and the firmware images:

1. Load the boot image with a non-staggered copy:

```
CN 4093(config)# copy {tftp|ftp|sftp} boot-image {address <IP address>}
{filename <image filename>}
```

2. Load the firmware image with a staggered copy:

```
CN 4093(config)# copy {tftp|ftp|sftp} {image1|image2} {address <IP address>}
{filename <image filename>} staggered-upgrade [delay <2-20 minutes>]
```

Saving Syslog Messages

By default, syslog messages on each member of a stack are saved to flash memory on that stack member. You may want to preserve stacking-related errors. To accomplish this, in console mode, use the following command:

```
CN 4093(config)# [no] logging log stacking
```

The master switch can display the syslog messages originated on any stack member as long as the specified stack element is currently an active member of the stack using the command:

```
CN 4093(config)# show logging [swn <configured-switch-number>] [messages|
|reverse|severity <0-7>]
```

where:

<*configured-switch-number*>	The configured switch number. If no number is supplied, the command applies to the master switch.
messages	shows last 2000 syslog messages.
reverse	shows syslog information in reverse priority order.
severity <*0-7*>	shows messages of a specific severity level.

For example, to retrieve the last 2000 syslog messages of severity 4 or greater from switch 3, enter:

```
CN 4093(config)# show logging swn 3 severity 4
```

To retrieve the contents of the log files stored on flash on a specified switch in the stack and copy that information to an external host using the specified protocol (SFTP or TFTP). In case the feature of saving log to flash is disabled, this command must be rejected.

To copy syslog content to an external host using SFTP or TFTP, use the command:

```
CN 4093(config)# copy log [swn <switch number>] {stfp|tftp [address <address>]
[filename <filename>]}
```

where:

<*switch number*>	The configured switch number. If no number is supplied, the command applies to the master switch.
address	The IP address of the TFTP host.
filename	The filename on the TFTP host.

For example:

```
CN 4093(config)# copy log tftp 192.168.1.85 // Copy logs from clients on
the master
```

```
CN 4093(config)# copy log swn 3 tftp 10.10.10.1 // Copy logs from stack
member 3
```

To configure up to two external hosts to log stack errors, use the command:

```
CN 4093(config)# logging host <host instance> {address <IPv4 address>|address6
<IPv6 address>|facility <facility (0-7)>|severity <severity (0-7)>}
```

where:

<host instance>	The host instance; either 1 or 2.
<IPv4 address>	The IPv4 address of the host being logged.
<IPv6 address>	The IPv6 address of the host being logged.
<facility (0-7)>	The facility (0-7) of the logs being written to external syslog servers.
<severity (0-7)>	The severity (0-7) of the logs being written to external syslog servers.

To enable console output of syslog messages, use the command:

```
CN 4093(config)# logging console severity <severity (0-7)>
```

where *<severity>* configures the severity of logs to be sent to the console.

To configure the severity of syslogs written to flash, use the command:

```
CN 4093(config)# logging buffer severity <severity (0-7)>
```

where *<severity>* configures the severity of logs to be written to flash.

Flexible Port Mapping in Stacking

Flexible Port Mapping allows administrators to manually enable or disable specific switch ports within the limitations of the installed licenses' bandwidth. For details, see "Flexible Port Mapping" on page 584.

In stacking, there are no overall bandwidth restrictions. Instead, each switch in the stack that supports licensing has bandwidth restrictions determined by its license level.

Commands associated with flexible port mapping can only be run from the master switch in the stack and can have an additional parameter:

- **`[no] boot port-map`** *<csnum:port number or range>*

  Adds or removes ports of a stack switch to/from the port map by specifying the switch's configured number and port number or range of ports. For example:

  ```
  CN 4093(config)# boot port-map 3:12
  ```
  Adds port number 12 of stack configured switch number 3 to the port map.

- **`default boot port-map [`***<csnum>***`]`**

  Resets the port map configuration to the default settings for the whole stack. If a configured switch number is specified, the command will reset the port map configuration only for the selected stack switch.

- **show boot port-map [**<*csnum*>**]**

  Displays the current port map configuration for the whole stack. If a configured switch number is specified, the command will display the port map configuration only for the selected stack switch.

```
CN 4093> show boot port-map


Switch 1
Maximum bandwidth: 640G
Used bandwidth: 640G
Mapped ports:
        1:1 1:2 1:3 1:4 1:5 1:6 1:7 1:8 1:9 1:10
        1:11 1:12 1:13 1:14 1:15 1:16 1:17 1:18 1:19 1:20
        1:21 1:22 1:23 1:24 1:25 1:26 1:27 1:28 1:29 1:30
        1:31 1:32 1:33 1:34 1:35 1:36 1:37 1:38 1:39 1:40
        1:41 1:42
        1:43 1:44 1:45 1:49 1:53 1:54 1:55 1:56 1:57 1:58
        1:59 1:60 1:61 1:62 1:63 1:64
Unmapped ports:


Switch 2
Maximum bandwidth: 640G
Used bandwidth: 640G
Mapped ports:
        2:1 2:2 2:3 2:4 2:5 2:6 2:7 2:8 2:9 2:10
        2:11 2:12 2:13 2:14 2:15 2:16 2:17 2:18 2:19 2:20
        2:21 2:22 2:23 2:24 2:25 2:26 2:27 2:28 2:29 2:30
        2:31 2:32 2:33 2:34 2:35 2:36 2:37 2:38 2:39 2:40
        2:41 2:42
        2:43 2:44 2:45 2:46 2:47 2:48 2:49 2:50 2:51 2:52
        2:53 2:54 2:55 2:56 2:57 2:61
Unmapped ports:

...
```

```
CN 4093> show boot port-map 3


Switch 3
Maximum bandwidth: 640G
Used bandwidth: 640G
Mapped ports:
        3:1 3:2 3:3 3:4 3:5 3:6 3:7 3:8 3:9 3:10
        3:11 3:12 3:13 3:14 3:15 3:16 3:17 3:18 3:19 3:20
        3:21 3:22 3:23 3:24 3:25 3:26 3:27 3:28 3:29 3:30
        3:31 3:32 3:33 3:34 3:35 3:36 3:37 3:38 3:39 3:40
        3:41 3:42
        3:43 3:44 3:45 3:49 3:53 3:54 3:55 3:56 3:57 3:58
        3:59 3:60 3:61 3:62 3:63 3:64
Unmapped ports:
```

ISCLI Stacking Commands

Stacking-related ISCLI commands are listed here. For details on specific commands, see the *CN4093 10 Gb Converged Scalable Switch Command Reference*.

- **[no] boot stack enable**
- **boot stack higig-trunk** *<port alias or number>*
- **boot stack mode {master|member} [***<asnum>***|master|backup|all]**
- **boot stack push-image {boot-image|image1|image2}** *<asnum>*
- **boot stack vlan** *<VLAN ID>*
- **connect** *<asnum>*
- **copy log [swn** *<switch number>***] stfp**
- **copy log [swn** *<switch number>***] tftp address** *<IP address>* **filename** *<file path>*
- **default boot stack [***<asnum>***|master|backup|all]**
- **logging buffer severity** *<severity (0-7)>*
- **logging console severity** *<severity (0-7)>*
- **logging host** *<host instance (1-2)>* **{address** *<IP address>***|facility** *<facility (0-7)>***|severity** *<severity (0-7)>*
- **[no] logging log stacking**
- **[no] stack backup** *<csnum>*
- **stack bind**
- **[no] stack name** *<1-63 characters>*
- **[no] stack switch-number** *<csnum>* **[description]**
- **show boot stack {***<asnum>***|master|backup|all}**
- **show logging [swn** *<csnum>***] [messages|reverse|severity** *<severity 0-7>***]**
- **show interface link switch** *<csnum>* **type [stacking| non-stacking]**
- **show interface link type [stacking|non-stacking]**
- **show interface link** *<port alias or number>* **type [stacking| non-stacking]**
- **show stack attached-switches**
- **show stack backup**
- **show stack dynamic**
- **show stack link**
- **show stack name**
- **show stack path-map [***<csnum>***]**
- **show stack push-status**
- **show stack switch**
- **show stack switch-number [***<csnum>***]**
- **show stack version**
- **stack backup** *<csnum>*
- **stack name** *<word>*
- **stack switch-number** *<csnum>* **bay** *<Slot ID>* **[universal-unic-id** *<uuid>***]**
- **stack switch-number** *<csnum>* **bind** *<asnum>*
- **stack switch-number** *<csnum>* **description** *<1-63 characters>*
- **stack switch-number** *<csnum>* **universal-unic-id** *<uuid>* **[bay** *<Slot ID>***]**

Chapter 14. Virtualization

Virtualization allows resources to be allocated in a fluid manner based on the logical needs of the data center, rather than on the strict, physical nature of components. The following virtualization features are included in Enterprise NOS 8.4 on the CN4093 10 Gb Converged Scalable Switch (CN4093):

- Virtual Local Area Networks (VLANs)

  VLANs are commonly used to split groups of networks into manageable broadcast domains, create logical segmentation of workgroups, and to enforce security policies among logical network segments.

  For details on this feature, see "VLANs" on page 141.

- Port aggregation

  A port LAG pools multiple physical switch ports into a single, high-bandwidth logical link to other devices. In addition to aggregating capacity, LAGs provides link redundancy.

  For details on this feature, see "Ports and Link Aggregation (LAG)" on page 161.

- Virtual Network Interface Card (vNIC) support

  Some NICs, such as the Emulex Virtual Fabric Adapter, can virtualize NIC resources, presenting multiple virtual NICs to the server's OS or hypervisor. Each vNIC appears as a regular, independent NIC with some portion of the physical NIC's overall bandwidth. ENOS 8.4 supports up to four vNICs over each internal switch port.

  For details on this feature, see "Virtual NICs" on page 265.

- Virtual Link Aggregation Groups (VLAGs)

  With VLAGs, two switches can act as a single logical device for the purpose of establishing port LAGs. Active LAG links from one device can lead to both VLAG peer switches, providing enhanced redundancy, including active-active VRRP configuration.

  For details on this feature, see "Virtual Link Aggregation Groups" on page 197

- VMready

  The switch's VMready software makes it *virtualization aware*. Servers that run hypervisor software with multiple instances of one or more operating systems can present each as an independent *virtual machine* (VM). With VMready, the switch automatically discovers virtual machines (VMs) connected to switch.

  For details on this feature, see "VMready" on page 283.

- Edge Virtual Bridging (QBG)

  The 802.1Qbg/Edge Virtual Bridging (EVB) is an emerging IEEE standard for allowing networks to become virtual machine (VM)-aware. EVB bridges the gap between physical and virtual network resources.

  For details on this feature, see "Edge Virtual Bridging" on page 353.

- Unified Fabric Port (UFP)

  An architecture that logically subdivides a high-speed physical link connecting to a server NIC or to a Converged Network Adapter (CNA). UFP provides a switch fabric component to control the NIC.

  For details on this feature, see "Unified Fabric Port" on page 365.

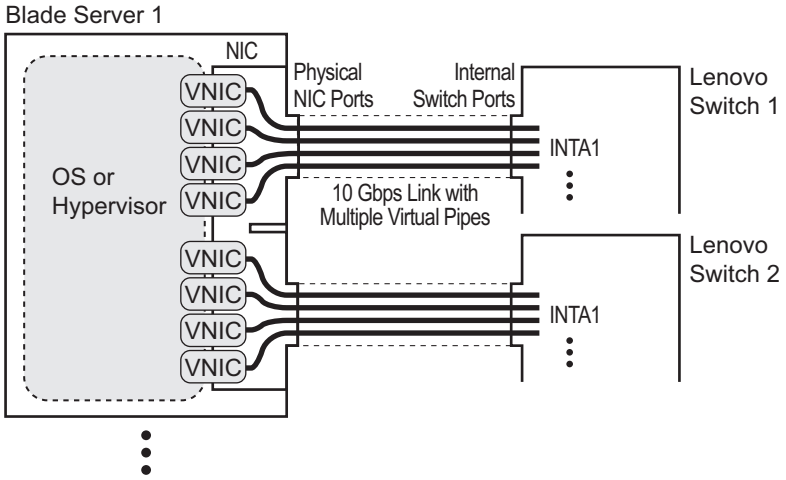
Enterprise NOS virtualization features provide a highly-flexible framework for allocating and managing switch resources.

Chapter 15. Virtual NICs

A Network Interface Controller (NIC) is a component within a blade server that allows the server to be connected to a network. The NIC provides the physical point of connection, as well as internal software for encoding and decoding network packets.

Virtualizing the NIC helps to resolve issues caused by limited NIC slot availability. By virtualizing a 10Gbps NIC, its resources can be divided into multiple logical instances known as virtual NICs (vNICs). Each vNIC appears as a regular, independent NIC to the server operating system or a hypervisor, with each vNIC using some portion of the physical NIC's overall bandwidth.

Figure 26. Virtualizing the NIC for Multiple Virtual Pipes on Each Link



A CN4093 with Enterprise NOS 8.4 supports the Emulex Virtual Fabric Adapter (VFA) 2-port 10Gb LOM and Emulex Virtual Fabric Adapter (Fabric Mezz) for Lenovo Flex System to provide the following vNIC features:

- Up to four vNICs are supported on each internal switch port.
- Each vNIC can accommodate one of the following traffic types: regular Ethernet, iSCSI or Fibre Channel over Ethernet (FCoE).
- vNICs with traffic of the same type can be grouped together, along with regular internal ports, external uplink ports and LAGs, to define vNIC groups for enforcing communication boundaries.
- In the case of a failure on the external uplink ports associated with a vNIC group, the switch can signal affected vNICs for failover while permitting other vNICs to continue operation.
- Each vNIC can be allocated a symmetric percentage of the 10Gbps bandwidth on the link (from NIC to switch, and from switch to NIC).
- The CN4093 can be used as the single point of vNIC configuration.

The following restriction applies to vNICs:

- vNICs are not supported simultaneously with VM groups (see "VMready" on page 283) on the same switch ports.

By default, vNICs are disabled. The administrator can enable vNICs and configure vNIC features on the switch using the standard management options such as the Enterprise NOS CLI, the ISCLI, and the Browser-based Interface (BBI).

To enable the vNIC feature on the switch, use the following command on the vNIC Configuration Menu:

```
CN 4093(config)# vnic enable
```

Note: The Emulex Virtual Fabric Adapter for Lenovo Flex System can also operate in Physical NIC (PNIC) mode, in which case vNIC features are non-applicable.

vNIC IDs on the Switch

Enterprise NOS 8.4 supports up to four vNICs attached to each internal switch port. Each vNIC is provided its own independent virtual pipe on the port.

On stand-alone (non-stacked) switches, each vNIC is identified by port and vNIC number:

<port number or alias>`.`*<vNIC pipe number (1-4)>*

For example:

INTA1.1, INTA1.2, INTA1.3, and INTA1.4 represent the vNICs on port INTA1.

INTA2.1, INTA2.2, INTA2.3, and INTA2.4 represent the vNICs on port INTA2, etc.

These vNIC IDs are used when adding vNICs to vNIC groups, and are shown in some configuration and information displays.

Whenever switches are stacked, the switch `csnum` ID is also required:

<switch csnum>`:`*<port number or alias>*`.`*<vNIC pipe number (1-4)>*

For example:

2:INTA`2.3` refers to port INTA2, vNIC 3, switch number 2.

Note: The configuration examples in this chapter depict stand-alone (non-stacked) port and vNIC designations.

vNIC Interface Names on the Server

When running in virtualization mode, the Emulex Virtual Fabric Adapter presents eight vNICs to the OS or hypervisor (four for each of the two physical NIC ports). Each vNIC is identified in the OS or hypervisor with a different PCIe function number (0-7). PCIe function numbers correlate to vNIC IDs on the switch as follows:

For Emulex Virtual Fabric Adapter 2-port 10Gb LOM:

Table 20. *vNIC ID Correlation*

PCIe Function ID	NIC Port	Switch Slot	vNIC Pipe	vNIC ID
0	0	Bay 1	1	INTA*x*.1
2	0	Bay 1	2	INTA*x*.2
4	0	Bay 1	3	INTA*x*.3
6	0	Bay 1	4	INTA*x*.4
1	1	Bay 2	1	INTA*x*.1
3	1	Bay 2	2	INTA*x*.2
5	1	Bay 2	3	INTA*x*.3
7	1	Bay 2	4	INTA*x*.4

For Emulex Virtual Fabric Adapter (Fabric Mezz), when replacing the LOM card:

Table 21. *vNIC ID Correlation*

PCIe Function ID	NIC Port	Switch Slot	vNIC Pipe	vNIC ID
First ASIC				
0	1	Bay 1	1	INTA*x*.1
2	1	Bay 1	2	INTA*x*.2
4	1	Bay 1	3	INTA*x*.3
6	1	Bay 1	4	INTA*x*.4
1	2	Bay 2	1	INTA*x*.1
3	2	Bay 2	2	INTA*x*.2
5	2	Bay 2	3	INTA*x*.3
7	2	Bay 2	4	INTA*x*.4

Table 22. *vNIC ID Correlation*

PCIe Function ID	NIC Port	Switch Slot	vNIC Pipe	vNIC ID
Second ASIC				
0	3	Bay 1	1	INTB*x*.1
2	3	Bay 1	2	INTB*x*.2
4	3	Bay 1	3	INTB*x*.3
6	3	Bay 1	4	INTB*x*.4
1	4	Bay 2	1	INTB*x*.1
3	4	Bay 2	2	INTB*x*.2
5	4	Bay 2	3	INTB*x*.3
7	4	Bay 2	4	INTB*x*.4

For Emulex Virtual Fabric Adapter (Fabric Mezz), when adding it with the LOM Card:

Table 23. *vNIC ID Correlation*

PCIe Function ID	NIC Port	Switch Slot	vNIC Pipe	vNIC ID
First ASIC				
0	1	Bay 3	1	INTA*x*.1
2	1	Bay 3	2	INTA*x*.2
4	1	Bay 3	3	INTA*x*.3
6	1	Bay 3	4	INTA*x*.4
1	2	Bay 4	1	INTA*x*.1
3	2	Bay 4	2	INTA*x*.2
5	2	Bay 4	3	INTA*x*.3
7	2	Bay 4	4	INTA*x*.4

Table 24. *vNIC ID Correlation*

PCIe Function ID	NIC Port	Switch Slot	vNIC Pipe	vNIC ID
Second ASIC				
0	3	Bay 3	1	INTB*x*.1
2	3	Bay 3	2	INTB*x*.2
4	3	Bay 3	3	INTB*x*.3

Table 24. *vNIC ID Correlation*

PCIe Function ID	NIC Port	Switch Slot	vNIC Pipe	vNIC ID
6	3	Bay 3	4	INTB*x*.4
1	4	Bay 4	1	INTB*x*.1
3	4	Bay 4	2	INTB*x*.2
5	4	Bay 4	3	INTB*x*.3
7	4	Bay 4	4	INTB*x*.4

In this, the *x* in the vNIC ID represents the internal switch port and its corresponding server node of the vNIC pipe. Each physical NIC port is connected to a different switch bay in the blade chassis.

vNIC Uplink Modes

The switch supports two modes for configuring the vNIC uplinks: dedicated mode and shared mode. The default is the dedicated mode. To enable the shared mode, enter the following command:

```
CN 4093(config)# vnic uplink-share
```

In the dedicated mode, only one vNIC group is assigned to an uplink port. This port can be a regular port or a LAG port. The NIC places an outer tag on the vNIC group packets. This outer tag contains the vNIC group VLAN. The uplink NIC strips off the outer tag before sending out the packet. For details, see "vNIC Groups in Dedicated Mode" on page 274.

In the shared mode, multiple vNIC groups can be assigned to an uplink port. This port can be a regular port or a LAG port. The vNIC groups share the uplink. You may assign a few vNIC groups to share an uplink and the other vNIC groups to have a single uplink each. In either case, the switch still operates in shared mode. As in the dedicated mode, the NIC places an outer tag on the vNIC group packets. This outer tag contains the vNIC group VLAN. The uplink NIC does not strip off the outer tag. The vNIC group tag defines the regular VLAN for the packet.This behavior is particularly useful in cases where the downstream server does not set any tag. Effectively, each vNIC group is a VLAN, which you can assign by configuring the VLAN to the vNIC group. You must enable the tag configuration on the uplink port. For details, see "vNIC Groups in Shared Mode" on page 275.

The following table compares the configurations of the two modes.

Table 25. *Comparison: Dedicated Mode vs. Shared Mode*

Configuration Area	Dedicated Mode	Shared Mode
Port	"tagpvid" must be disabled.	"tagpvid" is user configurable.
	"pvid" = vNIC group VLAN.	"pvid" is user configurable.
	"tag" is user configurable.	"tag" must be enabled.
	Port can be added only to the vNIC group VLAN.	Port can be added to multiple VLANs in addition to the vNIC group VLANs that are automatically configured.
	Inserts vNIC group VLAN in the outer tag of ingress packets.	Inserts regular VLAN in the outer tag. VLAN tags are passed to and received from the uplink switch similar to vNIC ports.
		To handle untagged packets, configure the pvid/native VLAN of the uplink port to one of the vNIC group VLANs, and disable "tagpvid".
VLAN	Add the port to a vNIC group VLAN and delete it from any other VLAN when the vNIC group VLAN is enabled.	Add the port to all vNIC group VLANs that are sharing the port. Do not remove it from any other VLAN.
	Delete the port from the vNIC group VLAN and add it back to the default VLAN 1 when the vNIC group is disabled/deleted or when the vNIC feature is globally disabled.	Remove the port from a vNIC group VLAN when the vNIC group is disabled/deleted. When the vNIC feature is globally disabled or the port is not added in any vNIC group, remove the port from all vNIC group VLANs and add it back to default VLAN 1 if no non-vNIC VLAN exists on the port.
	Do not add a port or LAG to multiple vNIC groups that are enabled.	Can add a port or LAG to multiple vNIC groups that are enabled.
	Do not configure additional VLANs on the uplink ports.	Can configure additional VLANs on the uplink ports.
STP	An uplink port can only be in one STG.	An uplink port can be in multiple STGs.
	When you add a port to a vNIC group, STP is automatically disabled.	When you add a port to a vNIC group, STP is automatically disabled.
	When you remove a port from a vNIC group, STP is automatically reset to factory default.	When you remove a port from a vNIC group, STP is automatically reset to factory default.
Failover	An uplink up/event can trigger the failover state change only of one vNIC group.	An uplink up/event can trigger the failover state change of multiple vNIC groups.

vNIC Bandwidth Metering

Enterprise NOS 8.4 supports bandwidth metering for vNIC traffic. By default, each of the four vNICs on any given port is allowed an equal share (25%) of NIC capacity when enabled. However, you may configure the percentage of available switch port bandwidth permitted to each vNIC.

vNIC bandwidth can be configured as a value from 1 to 100, with each unit representing 1% (or 100Mbps) of the 10Gbps link. By default, each vNICs enabled on a port is assigned 25 units (equal to 25% of the link, or 2.5Gbps). When traffic from the switch to the vNIC reaches its assigned bandwidth limit, the switch will drop packets egressing to the affected vNIC.

Note: Bandwidth metering drops excess packets when configured limits are reached. Consider using the ETS feature in applications where packet loss is not desirable (see "Enhanced Transmission Selection" on page 358).

To change the bandwidth allocation, use the following commands:

```
CN 4093(config)# vnic port <port alias or number> index <vNIC number (1-4)>
CN 4093(vnic-config)# bandwidth <allocated percentage>
```

Note: vNICs that are disabled are automatically allocated a bandwidth value of 0.

A combined maximum of 100 units can be allocated among vNIC pipes enabled for any specific port (bandwidth values for disabled pipes are not counted). If more than 100 units are assigned to enabled pipes, an error will be reported when attempting to apply the configuration.

The bandwidth metering configuration is automatically synchronized between the switch and vNICs for regular Ethernet and iSCSI traffic. Once configured on the switch, there is no need to manually configure vNIC bandwidth metering limits on the NIC.

Note: FCoE vNIC does not use egress metering. ETS and PFC must be enabled to ensure lossless transmission for FCoE traffic. ETS does traffic shaping. You can configure a minimum bandwidth for each traffic class. For example, 40% for FCoE priority 3, 60% for the Ethernet traffic. FCoE traffic gets 40% minimum guaranteed bandwidth. If the Ethernet traffic only consumes 30% bandwidth, then FCoE traffic can use 70%. If there is no other Ethernet traffic, then FCoE traffic can use 100%. The FCoE vNIC can use up to 100% of the bandwidth, with a minimum guaranteed bandwidth of 40%.

vNIC Groups

vNICs can be grouped together, along with internal and external switch ports and LAGs, into vNIC groups. Each vNIC group is essentially a separate virtual network within the switch. Elements within a vNIC group have a common logical function and can communicate with each other, while elements in different vNIC groups are separated.

Enterprise NOS 8.4 supports up to 32 independent vNIC groups. To enforce group boundaries, each vNIC group is assigned its own unique VLAN.

The VLAN configured for the vNIC group will be automatically assigned to member vNICs, ports and LAGs and should not be manually configured for those elements.

Note: Once a VLAN is assigned to a vNIC group, that VLAN is used only for vNIC purposes and is no longer available for other configuration. Likewise, any VLAN configured for regular purposes cannot be configured as a vNIC group VLAN.

The vNIC group rules are as follows:

- vNIC groups may have one or more vNIC members. However, any given vNIC can be a member of only one vNIC group.
- All vNICs on a given port must belong to different vNIC groups.
- Each vNIC group may contain traffic of one type only (either regular Ethernet, iSCSI or FCoE). Traffic of different types may not be mixed within any vNIC group.
- External ports that are part of a LAG may not be individually added to a vNIC group. Only one individual external port, one static LAG or one dynamic LAG (consisting of multiple external ports) may be added to any given vNIC group.
- In dedicated mode, for any internal ports, external port or port LAG group connected to regular (non-vNIC) devices:
  - These elements can be placed in only one vNIC group (they cannot be members of multiple vNIC groups).
  - Once added to a vNIC group, the PVID for the element is automatically set to use the vNIC group VLAN number, and PVID tagging on the element is automatically disabled.
- By default, STP is disabled on any external port added to a vNIC group. STP can be re-enabled on the port if desired.
- Because regular, inner VLAN IDs are ignored by the switch for traffic in vNIC groups, following rules and restrictions apply:
  - The inner VLAN tag may specify any VLAN ID in the full, supported range (1 to 4095) and may even duplicate outer vNIC group VLAN IDs. However, in the shared mode, inner VLAN tag and the vNIC group VLAN ID should be the same.
  - Per-VLAN IGMP snooping is not supported in vNIC groups.
  - The inner VLAN tag is not processed in any way in vNIC groups: The inner tag cannot be stripped or added on port egress, is not used to restrict multicast traffic, is not matched against ACL filters, and does not influence Layer 3 switching.

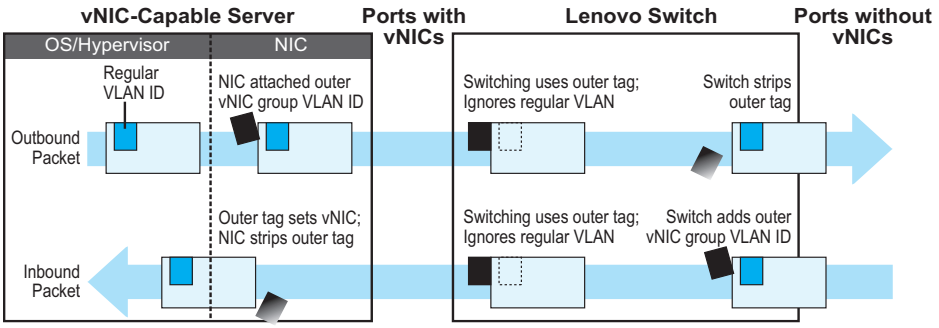
- For vNIC ports on the switch, because the outer vNIC group VLAN is transparent to the OS/hypervisor and upstream devices, VLAN tagging should be configured as normally required (on or off) for the those devices, ignoring any outer tag.

- Virtual machines (VMs) and other VEs associated with vNICs are automatically detected by the switch when VMready is enabled (see Chapter 16, "VMready"). However, vNIC groups are isolated from other switch elements. VEs in vNIC groups cannot be assigned to VM groups.

vNIC Groups in Dedicated Mode

The vNIC group VLAN ID is placed on all vNIC group packets as an "outer" tag. As shown in Figure 27, the outer vNIC group VLAN ID is placed on the packet in addition to any regular VLAN tag assigned by the network, server, or hypervisor. The outer vNIC group VLAN is used only between the CN4093 and the NIC.

Figure 27. Outer and Inner VLAN Tags



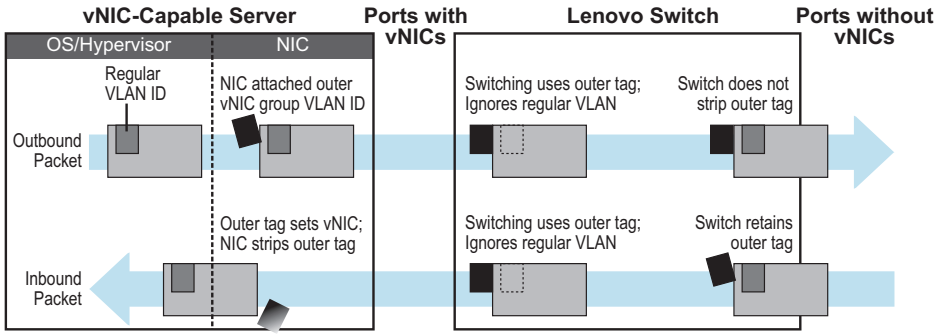
Within the CN4093, all Layer 2 switching for packets within a vNIC group is based on the outer vNIC group VLAN. The CN4093 does not consider the regular, inner VLAN ID (if any) for any VLAN-specific operation.

The outer vNIC group VLAN is removed by the NIC before the packet reaches the server OS or hypervisor, or by the switch before the packet egresses any internal port or external uplink port which does not need it for vNIC processing.

vNIC Groups in Shared Mode

The vNIC group VLAN ID is placed on all vNIC group packets as an “outer” tag. As shown in Figure 28, the outer vNIC group VLAN ID is placed on the packet in addition to any regular VLAN tag assigned by the network, server, or hypervisor.

Figure 28. Outer and Inner VLAN Tags



Within the CN4093, all Layer 2 switching for packets within a vNIC group is based on the outer vNIC group VLAN. The CN4093 does not consider the regular, inner VLAN ID (if any) for any VLAN-specific operation.

The outer vNIC group VLAN is not removed by the switch before the packet egresses any internal port or external uplink port. For untagged packets sent by the server, the uplink NIC uses this outer tag to switch the packet to destined VLAN.

The shared mode is useful is cases where the multiple vNIC groups need to share an uplink port. The vNIC group tag defines the user VLAN. Following is an use case:

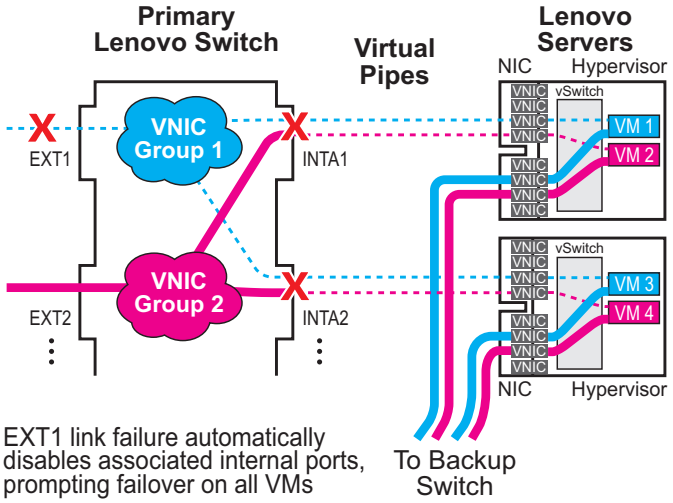
An ESX server is presented with eight vNICs (four from bay 7 and four from bay 9) used with four virtual switches of the ESX host and with no tagged port groups. A pair of odd/even vNICs is placed within each virtual switch. On the CN4093, four vNIC groups are created and the desired VLAN for each vNIC group is configured. For example, if vNIC group 1 on the CN4093 has four interfaces: 1.1, 2.1, 3.1, 4.1. vNIC group 1 is configured with VLAN 10. Packets coming from any VM connecting with the virtual switch that VMNIC 2 and 3 (vNIC 1.1, 2.1, 3.1, and 4.1 on bay 7 and bay 9) will be assigned with VLAN 10. These packets go out the uplink with VLAN 10 tag. The upstream switch sends these packets to the desired destination on VLAN 10.

vNIC Teaming Failover

For NIC failover in a non-virtualized environment, when a service group's external uplink ports fail or are disconnected, the switch disables the affected group's internal ports, causing the server to failover to the backup NIC and switch.

However, in a virtualized environment, disabling the affected internal ports would disrupt all vNIC pipes on those ports, not just those that have lost their external uplinks (see Figure 29).

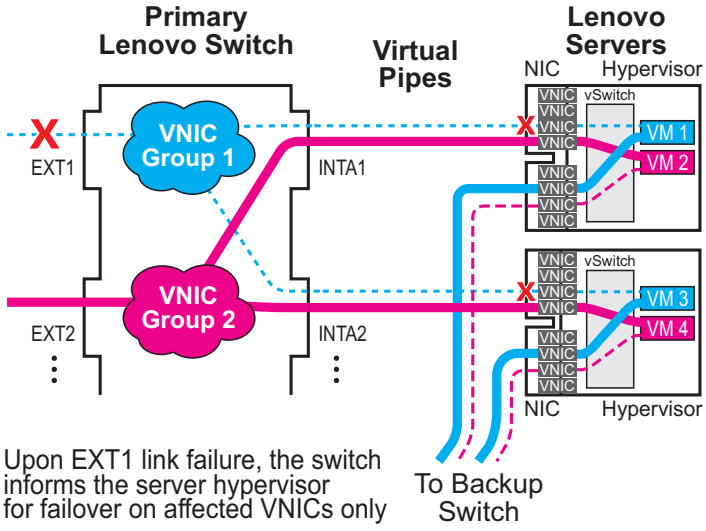
Figure 29. Regular Failover in a Virtualized Environment



To avoid disrupting vNICs that have not lost their external uplinks, ENOS 8.4 and the Emulex Virtual Fabric Adapter for Lenovo Flex System provide vNIC-aware failover.

In the dedicated mode, when a vNIC group's external uplink ports fail, the switch cooperates with the affected NIC to prompt failover only on the appropriate vNICs. This allows the vNICs that are not affected by the failure to continue without disruption (see Figure 30 on page 277).

Figure 30. vNIC Failover Solution



By default, vNIC Teaming Failover is disabled on each vNIC group, but can be enabled or disabled independently for each vNIC group using the following commands:

```
CN 4093(config)# vnic vnicgroup <group number>
CN 4093(vnic-group-config)# failover
```

vNIC Configuration Example

Consider the following example configuration of vNICs for regular Ethernet traffic:

Figure 31. Multiple vNIC Groups

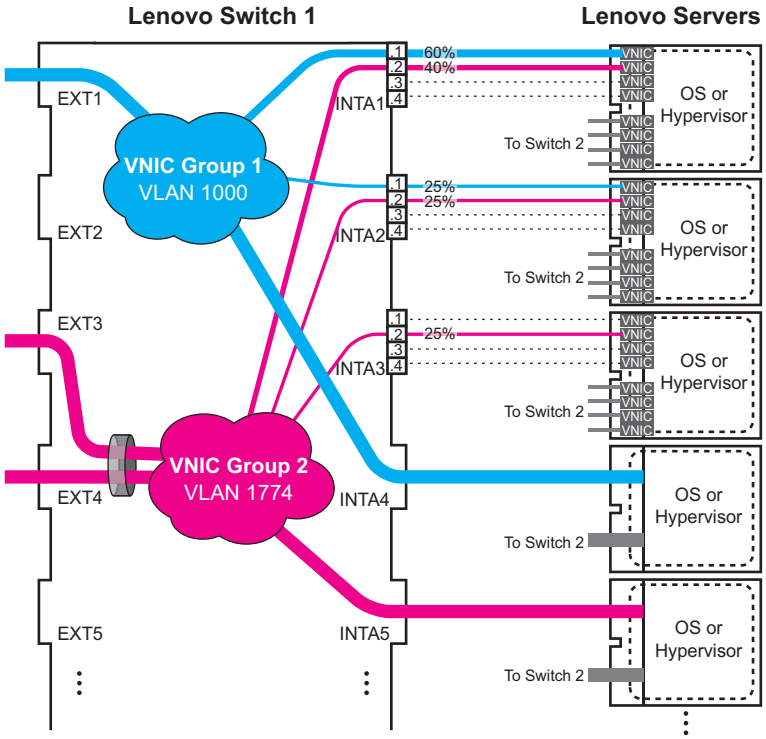


Figure 31 has the following vNIC network characteristics:

- vNIC group 1 has an outer tag for VLAN 1000. The group is comprised of vNIC pipes INTA1.1 and INTA2.1, internal port INTA4 (a non-vNIC port), and external uplink port EXT1.
- vNIC group 2 has an outer tag for VLAN 1774. The group is comprised of vNIC pipes INTA1.2, INTA2.2 and INTA3.2, internal port INTA5 and an external uplink LAG of ports EXT3 and EXT4.
- vNIC failover is enabled for both vNIC groups.
- vNIC bandwidth on port INTA1 is set to 60% for vNIC 1 and 40% for vNIC 2.
- Other enabled vNICs (INTA2.1, INTA2.2, and INTA3.2) are permitted the default bandwidth of 25% (2.5Gbsp) on their respective ports.
- All remaining vNICs are disabled (by default) and are automatically allocated 0 bandwidth.

1. Configure the external LAG to be used with vNIC group 2.

```
CN 4093(config)# portchannel 1 port EXT3,EXT4 enable
```

2. Enable the vNIC feature on the switch.

```
CN 4093(config)# vnic enable
```

3. Configure the virtual pipes for the vNICs attached to each internal port:

```
CN 4093(config)# vnic port INTA1 index 1      (Select vNIC 1 on the port)
CN 4093(vnic-config)# enable                   (Enable the vNIC pipe)
CN 4093(vnic-config)# bandwidth 60             (Allow 60% egress bandwidth)
CN 4093(vnic-config)# exit

CN 4093(config)# vnic port INTA1 index 2      (Select vNIC 2 on the port)
CN 4093(vnic-config)# enable                   (Enable the vNIC pipe)
CN 4093(vnic-config)# bandwidth 40             (Allow 40% egress bandwidth)
CN 4093(vnic-config)# exit

CN 4093(config)# vnic port INTA2 index 1      (Select vNIC 1 on the port)
CN 4093(vnic-config)# enable                   (Enable the vNIC pipe)
CN 4093(vnic-config)# exit

CN 4093(config)# vnic port INTA2 index 2      (Select vNIC 2 on the port)
CN 4093(vnic-config)# enable                   (Enable the vNIC pipe)
CN 4093(vnic-config)# exit
```

As a configuration shortcut, vNICs do not have to be explicitly enabled in this step. When a vNIC is added to the vNIC group (in the next step), the switch will prompt you to confirm automatically enabling the vNIC if it is not yet enabled (shown for INT3.2).

Note: vNICs are not supported simultaneously on the same switch ports as VMready.

4. Add ports, LAGs and virtual pipes to their vNIC groups.

```
CN 4093(config)# vnic vnicgroup 1                  (Select vNIC group)
CN 4093(vnic-group-config)# vlan 1000              (Specify the VLAN)
CN 4093(vnic-group-config)# member INTA1.1         (Add vNIC pipes to the group)
CN 4093(vnic-group-config)# member INTA2.1
CN 4093(vnic-group-config)# port INTA4             (Add non-vNIC port to the group)
CN 4093(vnic-group-config)# port EXT1              (Add uplink port to the group
CN 4093(vnic-group-config)# failover               (Enable vNIC failover for the group)
CN 4093(vnic-group-config)# enable                 (Enable the vNIC group)
CN 4093(vnic-group-config)# exit

CN 4093(config)# vnic vnicgroup 2
CN 4093(vnic-group-config)# vlan 1774
CN 4093(vnic-group-config)# member INTA1.2
CN 4093(vnic-group-config)# member INTA2.2
CN 4093(vnic-group-config)# member INTA3.2
vNIC 3.2 is not enabled.
Confirm enabling vNIC3.2 [y/n]: y
CN 4093(vnic-group-config)# port INTA5
CN 4093(vnic-group-config)# trunk 1
CN 4093(vnic-group-config)# failover
CN 4093(vnic-group-config)# enable
CN 4093(vnic-group-config)# exit
```

Once VLAN 1000 and 1774 are configured for vNIC groups, they will not be available for regular configuration.

Note: vNICs are not supported simultaneously on the same switch ports as VMready.

vNICs for iSCSI on Emulex Virtual Fabric Adapter

The ENOS vNIC feature works with standard network applications like iSCSI as previously described. However, the Emulex Virtual Fabric Adapter for Lenovo Flex System expects iSCSI traffic to occur only on a single vNIC pipe. When using the Emulex Adapter 2, only vNIC pipe 2 may participate in iSCSI.

To configure the switch for this solution, iSCSI traffic should be placed in its own vNIC group, comprised of the external uplink port leading to the iSCSI target, and the related <*port*>.2 vNIC pipes connected to the participating servers. For example:

1. Enable the vNIC feature on the switch.

```
CN 4093 # vnic enable
```

2. Configure the virtual pipes for the iSCSI vNICs attached to each internal port:

```
CN 4093(config)# vnic port INTA1 index 2      (Select vNIC 2 on the server port)
CN 4093(vnic_config)# enable                   (Enable the vNIC pipe)
CN 4093(vnic_config)# exit
CN 4093(config)# vnic port INTA2 index 2      (Select vNIC 2 on the server port)
CN 4093(vnic_config)# enable                   (Enable the vNIC pipe)
CN 4093(vnic_config)# exit
CN 4093(config)# vnic port INTA3 index 2      (Select vNIC 2 on the server port)
CN 4093(vnic_config)# enable                   (Enable the vNIC pipe)
CN 4093(vnic_config)# exit
```

Note: vNICs are not supported simultaneously on the same switch ports as VMready.

3. Add ports and virtual pipes to a vNIC group.

```
CN 4093(config)# vnic vnicgroup 1              (Select vNIC group)
CN 4093(vnic-group-config)# vlan 1000          (Specify the VLAN)
CN 4093(vnic-group-config)# member INTA1.2     (Add iSCSI vNIC pipes to the group)
CN 4093(vnic-group-config)# member INTA2.2
CN 4093(vnic-group-config)# member INTA3.2
CN 4093(vnic-group-config)# port EXT1          (Add the uplink port to the group)
CN 4093(vnic-group-config)# enable             (Enable the vNIC group)
CN 4093(vnic-group-config)# exit
```

vNICs for FCoE Using the Emulex VFA

Similar to the iSCSI application, when using the Emulex VFA for Lenovo chassis systems, FCoE traffic is expected to occur only on vNIC pipe 2. In this case, the additional vNIC configuration for FCoE support is minimal.

Consider an example where the Fibre Channel network is connected to an FCoE Forwarder (FCF) bridge via bridge port EXT4, and to an ENode on port INT1.

1. The following steps are required as part of the regular FCoE configuration (see "FIP Snooping Configuration" on page 353):
   a. Disable the FIP Snooping automatic VLAN creation.
   b. Disable FIP Snooping on all external ports not used for FCoE. FIP snooping should be enabled only on ports connected to an FCF or ENode.
   c. Turn on CEE and FIP Snooping.
   d. Manually configure the FCoE ports and VLAN: enable VLAN tagging on all FCoE ports, and place FCoE ports into a supported VLAN.

   When CEE is turned on and the regular FCoE configuration is complete, FCoE traffic will be automatically assigned to PFC priority 3, and be initially allocated 50% of port bandwidth via ETS.

   The following steps are specific to vNIC configuration.
2. On the NIC, ensure that FCoE traffic occurs on vNIC pipe 2 only. Refer to your Emulex VFA documentation for details.
3. On the switch, enable the vNIC feature.

```
CN 4093 # vnic enable
```

4. (Optional) For additional security, set the desired operation mode for FCoE ports:

```
CN 4093(config)# fcoe fips port INT1 fcf-mode off
(Select ENode port; Set as ENode connection)
CN 4093(config)# fcoe fips port EXT4 fcf-mode on
(Select FCF port; Set as FCF connection)
```

   No additional configuration for vNIC pipes or vNIC groups is required for FCoE. However, for other networks connected to the switch, appropriate vNIC pipes and vNIC groups should be configured as normal, if desired.

Chapter 16. VMready

Virtualization is used to allocate server resources based on logical needs, rather than on strict physical structure. With appropriate hardware and software support, servers can be virtualized to host multiple instances of operating systems, known as virtual machines (VMs). Each VM has its own presence on the network and runs its own service applications.

Software known as a *hypervisor* manages the various virtual entities (VEs) that reside on the host server: VMs, virtual switches, and so on. Depending on the virtualization solution, a virtualization management server may be used to configure and manage multiple hypervisors across the network. With some solutions, VMs can even migrate between host hypervisors, moving to different physical hosts while maintaining their virtual identity and services.

The Enterprise NOS 8.4 VMready feature supports up to 4096 VEs in a virtualized data center environment. The switch automatically discovers the VEs attached to switch ports, and distinguishes between regular VMs, Service Console Interfaces, and Kernel/Management Interfaces in a VMware® environment.

VEs may be placed into VM groups on the switch to define communication boundaries: VEs in the same VM group may communicate with each other, while VEs in different groups may not. VM groups also allow for configuring group-level settings such as virtualization policies and ACLs.

The administrator can also pre-provision VEs by adding their MAC addresses (or their IPv4 address or VM name in a VMware environment) to a VM group. When a VE with a pre-provisioned MAC address becomes connected to the switch, the switch will automatically apply the appropriate group membership configuration.

The CN4093 with VMready also detects the migration of VEs across different hypervisors. As VEs move, the CN4093 NMotion™ feature automatically moves the appropriate network configuration as well. NMotion gives the switch the ability to maintain assigned group membership and associated policies, even when a VE moves to a different port on the switch.

VMready also works with VMware Virtual Center (vCenter) management software. Connecting with a vCenter allows the CN4093 to collect information about more distant VEs, synchronize switch and VE configuration, and extend migration properties.

VE Capacity

When VMready is enabled, the switch will automatically discover VEs that reside in hypervisors directly connected on the switch ports. Enterprise NOS 8.4 supports up to 4096 VEs. Once this limit is reached, the switch will reject additional VEs.

Note: In rare situations, the switch may reject new VEs prior to reaching the supported limit. This can occur when the internal hash corresponding to the new VE is already in use. If this occurs, change the MAC address of the VE and retry the operation. The MAC address can usually be changed from the virtualization management server console (such as the VMware Virtual Center).

VM Group Types

VEs, as well as internal ports, external ports, static LAGs and LACP LAGs, can be placed into VM groups on the switch to define virtual communication boundaries. Elements in a given VM group are permitted to communicate with each other, while those in different groups are not. The elements within a VM group automatically share certain group-level settings.

Enterprise NOS 8.4 supports up to 4096 VM groups. There are two different types:

- Local VM groups are maintained locally on the switch. Their configuration is not synchronized with hypervisors.
- Distributed VM groups are automatically synchronized with a virtualization management server (see "Assigning a vCenter" on page 293).

Each VM group type is covered in detail in the following sections.

Local VM Groups

The configuration for local VM groups is maintained on the switch (locally) and is not directly synchronized with hypervisors. Local VM groups may include only local elements: local switch ports and LAGs, and only those VEs connected to one of the switch ports or pre-provisioned on the switch.

Local VM groups support limited VE migration: as VMs and other VEs move to different hypervisors connected to different ports on the switch, the configuration of their group identity and features moves with them. However, VE migration to and from more distant hypervisors (those not connected to the CN4093, may require manual configuration when using local VM groups).

Configuring a Local VM Group

Local VM groups are configured in the VM Group command path:

```
CN 4093(config)# virt vmgroup <VM group number>
```

Use the following ISCLI configuration commands to assign group properties and membership :

```
cpu                                                    (Enable sending unregistered IPMC to CPU)
flood                                                  (Enable flooding unregistered IPMC)
key <LACP LAG key>                                     (Add LACP LAG to group)
optflood                                               (Enable optimized flooding)
port <port alias or number>                            (Add port member to group)
portchannel <LAG group number>                         (Add static LAG to group)
profile <profile name>                                 (Not used for local groups)
stg <Spanning Tree group>                              (Add STG to group)
tag                                                    (Set VLAN tagging on ports)
validate <advanced|basic>                              (Validate mode for the group)
vlan <VLAN number>                                     (Specify the group VLAN)
vm <MAC>|<index>|<UUID>|<IPv4 address>|<name>          (Add VM member to group)
vmap <VMAP number> [intports|extports]                 (Specify VMAP number)
vport <Virtual port>                                   (Add a virtual port to the group)
```

The following rules apply to the local VM group configuration commands:

- `cpu`: Enable sending unregistered IPMC to CPU.
- `flood`: Enable flooding unregistered IPMC.
- `key`: Add LACP LAGs to the group.
- `optflood`: Enable optimized flooding to allow sending unregistered IPMC to the Mrouter ports without having any packet loss during the learning period; This option is disabled by default; When optflood is enabled, the flood and cpu settings are ignored.
- `port`: Add switch server ports or switch uplink ports to the group. Note that VM groups and vNICs (see "Virtual NICs" on page 265) are not supported simultaneously on the same port.
- `portchannel`: Add static port LAGs to the group.
- `profile`: The profile options are not applicable to local VM groups. Only distributed VM groups may use VM profiles (see "VM Profiles" on page 287).
- `stg`: The group may be assigned to a Spanning-Tree group for broadcast loop control (see"Spanning Tree Protocols" on page 175).
- `tag`: Enable VLAN tagging for the VM group. If the VM group contains ports which also exist in other VM groups, enable tagging in both VM groups.
- `validate`: Set validate mode for the group.
- `vlan`: Each VM group must have a unique VLAN number. This is required for local VM groups. If one is not explicitly configured, the switch will automatically assign the next unconfigured VLAN when a VE or port is added to the VM group.

- `vmap`: Each VM group may optionally be assigned a VLAN-based ACL (see "VLAN Maps" on page 296).
- `vm`: Add VMs.

  VMs and other VEs are primarily specified by MAC address. They can also be specified by UUID or by the index number as shown in various VMready information output (see "VMready Information Displays" on page 298).
- `vport`: Add a virtual port.

  Add a virtual port to the group.

Distributed VM Groups

Distributed VM groups allow configuration profiles to be synchronized between the CN4093 and associated hypervisors and VEs. This allows VE configuration to be centralized, and provides for more reliable VE migration across hypervisors.

Using distributed VM groups requires a virtualization management server. The management server acts as a central point of access to configure and maintain multiple hypervisors and their VEs (VMs, virtual switches, and so on).

The CN4093 must connect to a virtualization management server before distributed VM groups can be used. The switch uses this connection to collect configuration information about associated VEs, and can also automatically push configuration profiles to the virtualization management server, which in turn configures the hypervisors and VEs. See "Virtualization Management Servers" on page 293 for more information.

VM Profiles

VM profiles are required for configuring distributed VM groups. They are not used with local VM groups. A VM profile defines the VLAN and virtual switch bandwidth shaping characteristics for the distributed VM group. The switch distributes these settings to the virtualization management server, which in turn distributes them to the appropriate hypervisors for VE members associated with the group.

Creating VM profiles is a two part process. First, the VM profile is created as shown in the following command on the switch:

```
CN 4093(config)# virt vmprofile <profile name>
```

Next, the profile must be edited and configured using the following configuration commands:

```
CN 4093(config)# virt vmprofile edit <profile name> ?
  eshaping <average bandwidth> <burst size> <peak>
  shaping <average bandwidth> <burst size> <peak>
  vlan <VLAN number>
```

For virtual switch bandwidth shaping parameters, average and peak bandwidth are specified in kilobits per second (a value of 1000 represents 1 Mbps). Burst size is specified in kilobytes (a value of 1000 represents 1 MB).

Note: The bandwidth shaping parameters in the VM profile are used by the hypervisor virtual switch software. To set bandwidth policies for individual VEs, see "VM Policy Bandwidth Control" on page 297.

Once configured, the VM profile may be assigned to a distributed VM group as shown in the following section.

Initializing a Distributed VM Group

Note: A VM profile is required before a distributed VM group may be configured. See "VM Profiles" on page 287 for details.

Once a VM profile is available, a distributed VM group may be initialized using the following configuration command:

```
CN 4093(config)# virt vmgroup <VM group number> profile <VM profile name>
```

Only one VM profile can be assigned to a given distributed VM group. To change the VM profile, the old one must first be removed.

```
CN 4093(config)# no virt vmgroup <VM group number> profile
```

Note: The VM profile can be added only to an empty VM group (one that has no VLAN, VMs, or port members). Any VM group number currently configured for a local VM group (see "Local VM Groups" on page 284) cannot be converted and must be deleted before it can be used for a distributed VM group.

Assigning Members

VMs, ports and LAGs may be added to the distributed VM group only after the VM profile is assigned. Group members are added, pre-provisioned or removed from distributed VM groups in the same manner as with local VM groups ("Local VM Groups" on page 284), with the following exceptions:

- VMs: VMs and other VEs are not required to be local. Any VE known by the virtualization management server can be part of a distributed VM group.
- The VM group `vlan` option (see page 285) cannot be used with distributed VM groups. For distributed VM groups, the VLAN is assigned in the VM profile.

Synchronizing the Configuration

When the configuration for a distributed VM group is modified, the switch updates the assigned virtualization management server. The management server then distributes changes to the appropriate hypervisors.

For VM membership changes, hypervisors modify their internal virtual switch port groups, adding or removing internal port memberships to enforce the boundaries defined by the distributed VM groups. Virtual switch port groups created in this fashion can be identified in the virtual management server by the name of the VM profile, formatted as follows:

`Lenovo_`*<VM profile name>*

(or)

`Lenovo_`*<VM profile name>*`_`*<index number>* (for vDS profiles)

Using the VM Group command path (`CN 4093(config)#` **`virt vmgroup`** *<x>* *`vm`*) to add a server host interface to a distributed VM group does not create a new port group on the virtual switch or move the host. Instead, because the host interface already has its own virtual switch port group on the hypervisor, the VM profile settings are applied to its existing port group.

Note: When applying the distributed VM group configuration, the virtualization management server and associated hypervisors must take appropriate actions. If a hypervisor is unable to make requested changes, an error message will be displayed on the switch. Be sure to evaluate all error message and take the appropriate actions to be sure the expected changes are properly applied.

Removing Member VEs

Removing a VE from a distributed VM group on the switch will have the following effects on the hypervisor:

- The VE will be moved to the `Lenovo_Default` (to the `Lenovo_Default_`*<index number>* in case of vDS) port group in VLAN 0 (zero).
- Traffic shaping will be disabled for the VE.
- All other properties will be reset to default values inherited from the virtual switch.

VMcheck

The CN4093 primarily identifies virtual machines by their MAC addresses. An untrusted server or a VM could identify itself by a trusted MAC address leading to MAC spoofing attacks. Sometimes, MAC addresses get transferred to another VM, or they get duplicated.

The VMcheck solution addresses these security concerns by validating the MAC addresses assigned to VMs. The switch periodically sends hello messages on server ports. These messages include the switch identifier and port number. The hypervisor listens to these messages on physical NICs and stores the information, which can be retrieved using the VMware Infrastructure Application Programming Interface (VI API). This information is used to validate VM MAC addresses. Two modes of validation are available: Basic and Advanced.

Use the following command to select the validation mode or to disable validation:

```
CN 4093(config)# [no] virt vmgroup <VM group number> validate
{basic|advanced}
```

Basic Validation

This mode provides port-based validation by identifying the port used by a hypervisor. It is suitable for environments in which MAC reassignment or duplication cannot occur.

The switch, using the hello message information, identifies a hypervisor port. If the hypervisor port is found in the hello message information, it is deemed to be a trusted port. Basic validation should be enabled when:

- A VM is added to a VM group, and the MAC address of the VM interface is in the Layer 2 table of the switch.
- A VM interface that belongs to a VM group experiences a “source miss” i.e. is not able to learn new MAC address.
- A trusted port goes down. Port validation must be performed to ensure that the port does not get connected to an untrusted source when it comes back up.

Use the following command to set the action to be performed if the switch is unable to validate the VM MAC address:

```
CN 4093(config)# virt vmcheck action basic {log|link}

log - generates a log
link - disables the port
```

Advanced Validation

This mode provides VM-based validation by mapping a switch port to a VM MAC address. It is suitable for environments in which spoofing, MAC reassignment, or MAC duplication is possible.

When the switch receives frames from a VM, it first validates the VM interface based on the VM MAC address, VM Universally Unique Identifier (UUID), Switch port, and Switch ID available in the hello message information. Only if all the four parameters are matched, the VM MAC address is considered valid.

In advanced validation mode, if the VM MAC address validation fails, an ACL can be created to drop the traffic received from the VM MAC address on the switch port. Use the following command to specify the number of ACLs to be used for dropping traffic:

```
CN 4093(config)# virt vmcheck acls max <1-640>
```

Use the following command to set the action to be performed if the switch is unable to validate the VM MAC address:

```
CN 4093(config)# virt vmcheck action advanced {log|link|acl}
```

Following are the other VMcheck commands:

Table 26. *VMcheck Commands*

Command	Description
CN 4093(config)# **virt vmware hello {ena|hport** *<port number>***|haddr|htimer}**	Hello messages setting: enable/add port/advertise this IP address in the hello messages instead of the default management IP address/set the timer to send the hello messages
CN 4093(config)# **no virt vmware hello {enable|hport** *<port number>***}**	Disable hello messages/remove port
CN 4093(config)# **[no] virt vmcheck trust** *<port number>*	Mark a port as trusted; Use the no form of the command to mark port as untrusted
CN 4093# **no virt vmcheck acl [mac-address [***<port number>***]|port]**	Delete ACL(s): all ACLs/an ACL by MAC address ((optional) and port number) /all ACLs installed on a port

Virtual Distributed Switch

A virtual Distributed Switch (vDS) allows the hypervisor's NIC to be attached to the vDS instead of its own virtual switch. The vDS connects to the vCenter and spans across multiple hypervisors in a datacenter. The administrator can manage virtual machine networking for the entire data center from a single interface. The vDS enables centralized provisioning and administration of virtual machine networking in the data center using the VMware vCenter server.

When a member is added to a distributed VM group, a distributed port group is created on the vDS. The member is then added to the distributed port group.

Distributed port groups on a vDS are available to all hypervisors that are connected to the vDS. Members of a single distributed port group can communicate with each other.

Note: vDS works with ESX 4.0 or higher versions.

To add a vDS, use the command:

```
CN 4093# virt vmware dvswitch add <datacenter name> <dvSwitch name>
[<dvSwitch-version>]
```

Prerequisites

Before adding a vDS on the CN4093, ensure the following:

- VMware vCenter is fully installed and configured and includes a "`bladevm`" administration account and a valid SSL certificate.
- A virtual distributed switch instance has been created on the vCenter. The vDS version must be higher or the same as the hypervisor version on the hosts.
- At least two hypervisors are configured.

Guidelines

Before migrating VMs to a vDS, consider the following:

- At any one time, a VM NIC can be associated with only one virtual switch: to the hypervisor's virtual switch, or to the vDS.
- Management connection to the server must be ensured during the migration. The connection is via the Service Console or the Kernel/Management Interface.
- The vDS configuration and migration can be viewed in vCenter at the following locations:
  - vDS: `Home> Inventory > Networking`
  - vDS Hosts: `Home > Inventory > Networking > vDS > Hosts`

  Note: These changes will not be displayed in the running configuration on the CN4093.
- When adding or moving a VM to a VM profile on a VM host, the path between the ESX servers and the vCenter must be up before you load the configuration for the operation to be successful.

Migrating to vDS

You can migrate VMs to the vDS using vCenter. The migration may also be accomplished using the operational commands on the CN4093 available in the following CLI menus:

For VMware vDS operations:

```
CN 4093# virt vmware dvswitch ?
add           Add a dvSwitch to a DataCenter
addhost       Add a host to a dvSwitch
adduplnk      Add a physical NIC to dvSwitch uplink ports
del           Remove a dvSwitch from a DataCenter
remhost       Remove a host from a dvSwitch
remuplnk      Remove a physical NIC from dvSwitch uplink ports
```

For VMware distributed port group operations:

```
CN 4093# virt vmware dpg ?
add           Add a port group to a dvSwitch
del           Delete a port group from a dvSwitch
update        Update a port group on a dvSwitch
vmac          Change a VM NIC's port group
```

Virtualization Management Servers

The CN4093 can connect with a virtualization management server to collect configuration information about associated VEs. The switch can also automatically push VM group configuration profiles to the virtualization management server, which in turn configures the hypervisors and VEs, providing enhanced VE mobility.

One virtual management server must be assigned on the switch before distributed VM groups may be used. Enterprise NOS 8.4 currently supports only the VMware Virtual Center (vCenter).

Assigning a vCenter

Assigning a vCenter to the switch requires the following:

- The vCenter must have a valid IPv4 address which is accessible to the switch (IPv6 addressing is not supported for the vCenter).
- A user account must be configured on the vCenter to provide access for the switch. The account must have (at a minimum) the following vCenter user privileges:
  - Network
  - Host Network > Configuration
  - Virtual Machine > Modify Device Settings

Once vCenter requirements are met, the following configuration command can be used on the CN4093 to associate the vCenter with the switch:

```
CN 4093(config)# virt vmware vcspec <vCenter IPv4 address> <username> [noauth]
```

This command specifies the IPv4 address and account username that the switch will use for vCenter access. Once entered, the administrator will be prompted to enter the password for the specified vCenter account.

The `noauth` option causes to the switch to ignores SSL certificate authentication. This is required when no authoritative SSL certificate is installed on the vCenter.

Note: By default, the vCenter includes only a self-signed SSL certificate. If using the default certificate, the `noauth` option is required.

Once the vCenter configuration has been applied on the switch, the CN4093 will connect to the vCenter to collect VE information.

vCenter Scans

Once the vCenter is assigned, the switch will periodically scan the vCenter to collect basic information about all the VEs in the datacenter, and more detailed information about the local VEs that the switch has discovered attached to its own ports.

The switch completes a vCenter scan approximately every two minutes. Any major changes made through the vCenter may take up to two minutes to be reflected on the switch. However, you can force an immediate scan of the vCenter by using one of the following ISCLI privileged EXEC commands:

```
CN 4093# virt vmware scan              (Scan the vCenter)
  -or-
CN 4093# show virt vm -v -r            (Scan vCenter and display result)
```

Deleting the vCenter

To detach the vCenter from the switch, use the following configuration command:

```
CN 4093(config)# no virt vmware vcspec
```

Note: Without a valid vCenter assigned on the switch, any VE configuration changes must be manually synchronized.

Deleting the assigned vCenter prevents synchronizing the configuration between the CN4093 and VEs. VEs already operating in distributed VM groups will continue to function as configured, but any changes made to any VM profile or distributed VM group on the switch will affect only switch operation; changes on the switch will not be reflected in the vCenter or on the VEs. Likewise, any changes made to VE configuration on the vCenter will no longer be reflected on the switch.

Exporting Profiles

VM profiles for discovered VEs in distributed VM groups are automatically synchronized with the virtual management server and the appropriate hypervisors. However, VM profiles can also be manually exported to specific hosts before individual VEs are defined on them.

By exporting VM profiles to a specific host, BNT port groups will be available to the host's internal virtual switches so that new VMs may be configured to use them.

VM migration requires that the target hypervisor includes all the virtual switch port groups to which the VM connects on the source hypervisor. The VM profile export feature can be used to distribute the associated port groups to all the potential hosts for a given VM.

A VM profile can be exported to a host using the following ISCLI privileged EXEC command:

```
CN 4093# virt vmware export <VM profile name> <host list> <virtual switch name>
```

The host list can include one or more target hosts, specified by host name, IPv4 address, or UUID, with each list item separated by a space. If the virtual switch name is omitted, the administrator will be prompted to select one from a list or to enter a new virtual switch name.

Once executed, the requisite port group will be created on the specified virtual switch. If the specified virtual switch does not exist on the target host, it will be created with default properties, but with no uplink connection to a physical NIC (the administrator must assign uplinks using VMware management tools.

VMware Operational Commands

The CN4093 may be used as a central point of configuration for VMware virtual switches and port groups using the VMware operational menu, available with the following ISCLI privileged EXEC commands:

```
CN 4093# virt vmware ?
dpg         Distributed port group operations
dvswitch    VMWare dvSwitch operations
export      Create or update a vm profile on one host
pg          Add a port group to a host
scan        Perform a VM Agent scan operation now
updpg       Update a port group on a host
vmacpg      Change a vnic's port group
vsw         Add a vswitch to a host
```

Pre-Provisioning VEs

VEs may be manually added to VM groups in advance of being detected on the switch ports. By pre-provisioning the MAC address of VEs that are not yet active, the switch will be able to later recognize the VE when it becomes active on a switch port, and immediately assign the proper VM group properties without further configuration.

Undiscovered VEs are added to or removed from VM groups using the following configuration commands:

```
CN 4093(config)# [no] virt vmgroup <VM group number> vm <VE MAC address>
```

For the pre-provisioning of undiscovered VEs, a MAC address is required. Other identifying properties, such as IPv4 address or VM name permitted for known VEs, cannot be used for pre-provisioning.

VLAN Maps

A VLAN map (VMAP) is a type of Access Control List (ACL) that is applied to a VLAN or VM group rather than to a switch port as with regular ACLs (see "Access Control Lists" on page 125). In a virtualized environment, VMAPs allow you to create traffic filtering and metering policies that are associated with a VM group VLAN, allowing filters to follow VMs as they migrate between hypervisors.

VMAPs are configured using the following ISCLI configuration command path:

```
CN 4093(config)# access-control vmap <VMAP ID> ?
  action         Set filter action
  egress-port    Set to filter for packets egressing this port
  ethernet       Ethernet header options
  ipv4           IP version 4 header options
  meter          ACL metering configuration
  mirror         Mirror options
  packet-format  Set to filter specific packet format types
  re-mark        ACL re-mark configuration
  statistics     Enable access control list statistics
  tcp-udp        TCP and UDP filtering options
```

Enterprise NOS 8.4 supports up to 128 VMAPs. Individual VMAP filters are configured in the same fashion as regular ACLs, except that VLANs cannot be specified as a filtering criteria (unnecessary, since VMAPs are assigned to a specific VLAN or associated with a VM group VLAN).

Once a VMAP filter is created, it can be assigned or removed using the following commands:

- For regular VLANs, use config-vlan mode:

```
CN 4093(config)# vlan <VLAN ID>
CN 4093(config-vlan)# [no] vmap <VMAP ID> [intports| extports]
```

- For a VM group, use the global configuration mode:

```
CN 4093(config)# [no] virt vmgroup <ID> vmap <VMAP ID>
[intports|extports]
```

Note: Each VMAP can be assigned to only one VLAN or VM group. However, each VLAN or VM group may have multiple VMAPs assigned to it.

The optional `intports` or `extports` parameter can be specified to apply the action (to add or remove the VMAP) for either the internal ports or external ports only. If omitted, the operation will be applied to all ports in the associated VLAN or VM group.

Note: VMAPs have a lower priority than port-based ACLs. If both an ACL and a VMAP match a particular packet, both filter actions will be applied as long as there is no conflict. In the event of a conflict, the port ACL will take priority, though switch statistics will count matches for both the ACL and VMAP.

VM Policy Bandwidth Control

In a virtualized environment where VEs can migrate between hypervisors and thus move among different ports on the switch, traffic bandwidth policies must be attached to VEs, rather than to a specific switch port.

VM Policy Bandwidth Control allows the administrator to specify the amount of data the switch will permit to flow to or from a particular VE, without defining a complicated matrix of ACLs or VMAPs for all port combinations where a VE may appear.

VM Policy Bandwidth Control Commands

VM Policy Bandwidth Control can be configured using the following configuration commands:

```
CN 4093(config)# virt vmpolicy vmbwidth <VM MAC>|<index>|<UUID>|
<IPv4 address>|<name> ?
  txrate <committed rate> <burst> [<ACL number>]     (Set the VM to switch rate)
  rxrate <committed rate> <burst>                    (Set the VM received bandwidth)
  bwctrl                                             (Enable bandwidth control)
```

Bandwidth allocation can be defined either for transmit (TX) traffic or receive (RX) traffic. Because bandwidth allocation is specified from the perspective of the VE, the switch command for TX Rate Control (`txrate`) sets the data rate to be sent from the VM to the switch, and the RX Rate Control (`rxrate`) sets the data rate to be received by the VM from the switch.

The *committed rate* is specified in multiples of 64 kbps, from 64 to 40,000,000. The maximum *burst* rate is specified as 32, 64, 128, 256, 1024, 2048, or 4096 kb. If both the committed rate and burst are set to 0, bandwidth control in that direction (TX or RX) will be disabled.

When `txrate` is specified, the switch automatically selects an available ACL for internal use with bandwidth control. Optionally, if automatic ACL selection is not desired, a specific ACL may be selected. If there are no unassigned ACLs available, `txrate` cannot be configured.

Bandwidth Policies vs. Bandwidth Shaping

VM Profile Bandwidth Shaping differs from VM Policy Bandwidth Control.

VM Profile Bandwidth Shaping (see "VM Profiles" on page 287) is configured per VM group and is enforced on the server by a virtual switch in the hypervisor. Shaping is unidirectional and limits traffic transmitted from the virtual switch to the CN4093. Shaping is performed prior to transmit VM Policy Bandwidth Control. If the egress traffic for a virtual switch port group exceeds shaping parameters, the traffic is dropped by the virtual switch in the hypervisor. Shaping uses server CPU resources, but prevents extra traffic from consuming bandwidth between the server and the CN4093.

VM Policy Bandwidth Control is configured per VE, and can be set independently for transmit and receive traffic. Bandwidth policies are enforced by the CN4093. VE traffic that exceeds configured levels is dropped by the switch upon ingress (for `txrate`) or before egress (for `rxrate`). Setting `txrate` uses ACL resources on the switch.

Bandwidth shaping and bandwidth policies can be used separately or in concert.

VMready Information Displays

The CN4093 can be used to display a variety of VMready information.

Note: Some displays depict information collected from scans of a VMware vCenter and may not be available without a valid vCenter. If a vCenter is assigned (see "Assigning a vCenter" on page 293), scan information might not be available for up to two minutes after the switch boots or when VMready is first enabled. Also, any major changes made through the vCenter may take up to two minutes to be reflected on the switch unless you force an immediate vCenter scan (see "vCenter Scans" on page 294).

Local VE Information

A concise list of local VEs and pre-provisioned VEs is available with the following ISCLI privileged EXEC command:

```
CN 4093# show virt vm
Virtual MAC addresses information:
IP Address  VMAC Address       Index  Port     VM Group (Profile)  Check status
----------  -----------------  -----  -------  ------------------  ------------
0.0.0.0     00:00:00:00:11:11     0   UNKNOWN       100
VMReady ports:  INTA1-INTC14

Number of entries: 1
0.0.0.0 indicates IP address not yet available

EVB Virtual Station Interface Information:
Total number of VM Association entries : 0
```

Note: The Index numbers shown in the VE information displays can be used to specify a particular VE in configuration commands.

If a vCenter is available, more verbose information can be obtained using the following ISCLI privileged EXEC command:

```
CN 4093# show virt vm -v

Index  MAC Address,       Name (VM or Host),  Port,  Group  Vswitch,
       IP Address         @Host (VMs only)    VLAN          Port Group
-----  ------------       ------------------  -----  -----  ----------
 0     00:50:56:9c:21:2f  atom                  4           vSwitch0
       172.16.46.15       @172.16.46.10       500            Eng_A

+1     00:50:56:72:ec:86  172.16.46.50          3           vSwitch0
       172.16.46.51                           0              VMkernel

*2     00:50:56:4f:f2:85  172.16.46.10          4           vSwitch0
       172.16.46.10                           0              Mgmt

+3     00:50:56:7c:1c:ca  172.16.46.10          4           vSwitch0
       172.16.46.11                           0              VMkernel

*4     00:50:56:4e:62:f5  172.16.46.50          3           vSwitch0
       172.16.46.50                           0              Mgmt

 5     00:50:56:9c:00:c8  quark                 4           vSwitch0
       172.16.46.25       @172.16.46.10       0              Corp

 6     00:50:56:9c:29:29  particle              3           vSwitch0
       172.16.46.35       @172.16.46.50       0              VM Network

 7     00:50:56:9c:47:fd  nucleus               3           vSwitch0
       172.16.46.45       @172.16.46.50       0              Finance

--
12 of 12 entries printed
* indicates VMware ESX Service Console Interface
+ indicates VMware ESX/ESXi VMkernel or Management Interface
```

To view additional detail regarding any specific VE, see "vCenter VE Details" on page 301).

vCenter Hypervisor Hosts

If a vCenter is available, the following ISCLI privileged EXEC command displays the name and UUID of all VMware hosts, providing an essential overview of the data center:

```
CN 4093# show virt vmware hosts
UUID                                 Name(s), IP Address
------------------------------------------------------------------
00a42681-d0e5-5910-a0bf-bd23bd3f7800  172.16.41.30
002e063c-153c-dd11-8b32-a78dd1909a00  172.16.46.10
00f1fe30-143c-dd11-84f2-a8ba2cd7ae00  172.16.44.50
0018938e-143c-dd11-9f7a-d8defa4b8300  172.16.46.20
...
```

Using the following command, the administrator can view more detailed vCenter host information, including a list of virtual switches and their port groups, as well as details for all associated VEs:

```
CN 4093# show virt vmware showhost {<UUID>|<IPv4 address>|<host name>}
Vswitches available on the host:
              vSwitch0
Port Groups and their Vswitches on the host:
              BNT_Default                  vSwitch0
              VM Network                   vSwitch0
              Service Console              vSwitch0
              VMkernel                     vSwitch0
------------------------------------------------------------------------
MAC Address         00:50:56:9c:21:2f
Port                4
Type                Virtual Machine
VM vCenter Name     halibut
VM OS hostname      localhost.localdomain
VM IP Address       172.16.46.15
VM UUID             001c41f3-ccd8-94bb-1b94-6b94b03b9200
Current VM Host     172.16.46.10
Vswitch             vSwitch0
Port Group          BNT_Default
VLAN ID             0
...
```

vCenter VEs

If a vCenter is available, the following ISCLI privileged EXEC command displays a list of all known VEs:

```
CN 4093# show virt vmware vms
UUID                                 Name(s), IP Address
------------------------------------------------------------------------
001cdf1d-863a-fa5e-58c0-d197ed3e3300  30vm1
001c1fba-5483-863f-de04-4953b5caa700  VM90
001c0441-c9ed-184c-7030-d6a6bc9b4d00  VM91
001cc06e-393b-a36b-2da9-c71098d9a700  vm_new
001c6384-f764-983c-83e3-e94fc78f2c00  sturgeon
001c7434-6bf9-52bd-c48c-a410da0c2300  VM70
001cad78-8a3c-9cbe-35f6-59ca5f392500  VM60
001cf762-a577-f42a-c6ea-090216c11800  30VM6
001c41f3-ccd8-94bb-1b94-6b94b03b9200  halibut, localhost.localdomain,
                                      172.16.46.15
001cf17b-5581-ea80-c22c-3236b89ee900  30vm5
001c4312-a145-bf44-7edd-49b7a2fc3800  vm3
001caf40-a40a-de6f-7b44-9c496f123b00  30VM7
```

vCenter VE Details

If a vCenter is available, the following ISCLI privileged EXEC command displays detailed information about a specific VE:

```
CN 4093# show virt vmware showvm {<VM UUID>|<VM IPv4 address>|<VM name>}
------------------------------------------------------------------------
MAC Address         00:50:56:9c:21:2f
Port                4
Type                Virtual Machine
VM vCenter Name     halibut
VM OS hostname      localhost.localdomain
VM IP Address       172.16.46.15
VM UUID             001c41f3-ccd8-94bb-1b94-6b94b03b9200
Current VM Host     172.16.46.10
Vswitch             vSwitch0
Port Group          BNT_Default
VLAN ID             0
```

VMready Configuration Example

This example has the following characteristics:

- A VMware vCenter is fully installed and configured prior to VMready configuration and includes a "`bladevm`" administration account and a valid SSL certificate.
- The distributed VM group model is used.
- The VM profile named "Finance" is configured for VLAN 30, and specifies NIC-to-switch bandwidth shaping for 1Mbps average bandwidth, 2MB bursts, and 3Mbps maximum bandwidth.
- The VM group includes four discovered VMs on internal switch ports INT1A and INT2A, and one static LAG (previously configured) that includes external ports EXT2 and EXT2.

1. Enable the VMready feature.

```
CN 4093(config)# virt enable
```

2. Specify the VMware vCenter IPv4 address.

```
CN 4093(config)# virt vmware vcspec 172.16.100.1 bradevm
```

When prompted, enter the user password that the switch must use for access to the vCenter.

3. Create the VM profile.

```
CN 4093(config)# virt vmprofile Finance
CN 4093(config)# virt vmprofile edit Finance vlan 30
CN 4093(config)# virt vmprofile edit Finance shaping 1000 2000 3000
```

4. Define the VM group.

```
CN 4093(config)# virt vmgroup 1 profile Finance
CN 4093(config)# virt vmgroup 1 vm arctic
CN 4093(config)# virt vmgroup 1 vm monster
CN 4093(config)# virt vmgroup 1 vm sierra
CN 4093(config)# virt vmgroup 1 vm 00:50:56:4f:f2:00
CN 4093(config)# virt vmgroup 1 portchannel 1
```

When VMs are added, the internal server ports on which they appear are automatically added to the VM group. In this example, there is no need to manually add ports EXT1 and EXT2.

5. If necessary, enable VLAN tagging for the VM group:

```
CN 4093(config)# virt vmgroup 1 tag
```

Note: If the VM group contains ports which also exist in other VM groups, tagging should be enabled in both VM groups. In this example configuration, no ports exist in more than VM group.

Chapter 17. FCoE and CEE

This chapter provides conceptual background and configuration examples for using Converged Enhanced Ethernet (CEE) features of the CN4093 10 Gb Converged Scalable Switch, with an emphasis on Fibre Channel over Ethernet (FCoE) solutions. The following topics are addressed in this chapter:

- "Fibre Channel over Ethernet" on page 304

  Fibre Channel over Ethernet (FCoE) allows Fibre Channel traffic to be transported over Ethernet links. This provides an evolutionary approach toward network consolidation, allowing Fibre Channel equipment and tools to be retained, while leveraging cheap, ubiquitous Ethernet networks for growth.

  - "FCoE Initialization Protocol Snooping" on page 309

    Using FCoE Initialization Protocol (FIP) snooping, the CN4093 examines the FIP frames exchanged between ENodes and FCFs. This information is used to dynamically determine the ACLs required to block certain types of undesired or unvalidated traffic on FCoE links.

- "Converged Enhanced Ethernet" on page 306

  Converged Enhanced Ethernet (CEE) refers to a set of IEEE standards developed primarily to enable FCoE, requiring enhancing the existing Ethernet standards to make them lossless on a per-priority traffic basis, and providing a mechanism to carry converged (LAN/SAN/IPC) traffic on a single physical link. CEE features can also be utilized in traditional LAN (non-FCoE) networks to provide lossless guarantees on a per-priority basis, and to provide efficient bandwidth allocation.

  - "Priority-Based Flow Control" on page 316

    Priority-Based Flow Control (PFC) extends 802.3x standard flow control to allow the switch to pause traffic based on the 802.1p priority value in each packet's VLAN tag. PFC is vital for FCoE environments, where SAN traffic must remain lossless and must be paused during congestion, while LAN traffic on the same links is delivered with "best effort" characteristics.

  - "Enhanced Transmission Selection" on page 320

    Enhanced Transmission Selection (ETS) provides a method for allocating link bandwidth based on the 802.1p priority value in each packet's VLAN tag. Using ETS, different types of traffic (such as LAN, SAN, and management) that are sensitive to different handling criteria can be configured either for specific bandwidth characteristics, low-latency, or best-effort transmission, despite sharing converged links as in an FCoE environment.

  - "Data Center Bridging Capability Exchange" on page 326

    Data Center Bridging Capability Exchange Protocol (DCBX) allows neighboring network devices to exchange information about their capabilities. This is used between CEE-capable devices for the purpose of discovering their peers, negotiating peer configurations, and detecting misconfigurations.

Fibre Channel over Ethernet

Fibre Channel over Ethernet (FCoE) is an effort to converge two of the different physical networks in today's data centers. It allows Fibre Channel traffic (such as that commonly used in Storage Area Networks, or SANs) to be transported without loss over 10Gb Ethernet links (typically used for high-speed Local Area Networks, or LANs). This provides an evolutionary approach toward network consolidation, allowing Fibre Channel equipment and tools to be retained, while leveraging cheap, ubiquitous Ethernet networks for growth.

With server virtualization, servers capable of hosting both Fibre Channel and Ethernet applications will provide advantages in server efficiency, particularly as FCoE-enabled network adapters provide consolidated SAN and LAN traffic capabilities.

The Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch with Enterprise NOS 8.4 software is compliant with the INCITS T11.3, FC-BB-5 FCoE specification, supporting up to 2048 FCoE connections.

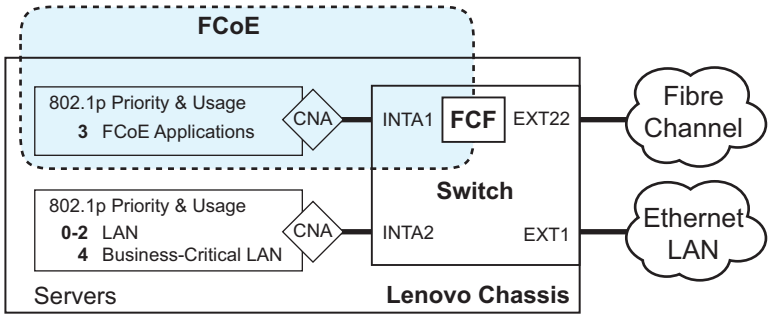
Note: Up to 2048 FCoE login sessions are supported. However, to achieve this using the internal FCF module, you need eight Omni ports (belonging to a single VLAN) connected to the FCF bridge.

The FCoE Topology

In an end-to-end Fibre Channel network, switches and end devices generally establish trusted, point-to-point links. Fibre Channel switches validate end devices, enforce zoning configurations and device addressing, and prevent certain types of errors and attacks on the network.

In a converged multi-hop FCoE network where Fibre Channel devices are bridged to Ethernet devices, the direct point-to-point QoS capabilities normally provided by the Fibre Channel fabric may be lost in the transition between the different network types. The CN4093 provides a solution to overcome this.

Figure 32. A Mixed Fibre Channel and FCoE Network



In Figure 32 on page 304, the FCoE network is connected to the Fibre Channel network through an FCoE Forwarder (FCF). The FCF acts as a Fibre Channel gateway to and from the multi-hop FCoE network. A full-fabric FC/FCoE switch or a Fibre Channel Node Port Virtualized (NPV) switch may perform the FCF function. Although it may be possible to use an external FCF device, this chapter focuses on using the built-in Fibre Channel features of the CN4093 itself.

For the FCoE portion of the network, the internal FCF is connected to a blade server (running Fibre Channel applications) through an FCoE-enabled Converged Network Adapter (CNA) known in Fibre Channel as an Ethernet Node (ENode).

Note: The figure also shows a non-FCoE LAN server connected to the CN4093 using a CNA. This allows the LAN server to take advantage of some CEE features that are useful even outside of an FCoE environment.

In order to block undesired or unvalidated traffic on FCoE links that exists outside the regular Fibre Channel topology, Ethernet ports used in FCoE are configured with Access Control Lists (ACLs) that are narrowly tailored to permit expected FCoE traffic to and from confirmed FCFs and ENodes, and deny all other FCoE or FIP traffic. This ensures that all FCoE traffic to an from the ENode passes through the FCF.

Because manual ACL configuration is an administratively complex task, the CN4093 can automatically and dynamically configure the ACLs required for use with FCoE. Using FCoE Initialization Protocol (FIP) snooping (see "FCoE Initialization Protocol Snooping" on page 309), the CN4093 examines the FIP frames normally exchanged between the FCF and ENodes to determine information about connected FCoE devices. This information is used to automatically determine the appropriate ACLs required to block certain types of undesired or unvalidated FCoE traffic.

Automatic FCoE-related ACLs are independent from ACLs used for typical Ethernet purposes.

FCoE Requirements

The following are required for implementing FCoE using the Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch (CN4093) with ENOS 8.4 software:

- The CN4093 must be connected to the Fibre Channel network using the switch's built-in Fibre Channel features (see "Fibre Channel" on page 333), or an external FCF device such as another CN4093 switch or a Lenovo RackSwitch G8264CS.
- For each CN4093 internal port participating in FCoE, the connected server must use the supported Converged Network Adapter (CNA) and must have the FCoE license enabled (if applicable) on the CNA.
- CEE must be turned on (see "Turning CEE On or Off" on page 306). When CEE is on, the DCBX, PFC, and ETS features are enabled and configured with default FCoE settings. These features may be reconfigured, but must remain enabled in order for FCoE to function.
- FIP snooping must be turned on (see "FCoE Initialization Protocol Snooping" on page 309). When FIP snooping is turned on, the feature is enabled on all ports by default. The administrator can disable FIP snooping on individual ports that do not require FCoE, but FIP snooping must remain enabled on all FCoE ports in order for FCoE to function.

Converged Enhanced Ethernet

Converged Enhanced Ethernet (CEE) refers to a set of IEEE standards designed to allow different physical networks with different data handling requirements to be converged together, simplifying management, increasing efficiency and utilization, and leveraging legacy investments without sacrificing evolutionary growth.

CEE standards were developed primarily to enable Fibre Channel traffic to be carried over Ethernet networks. This required enhancing the existing Ethernet standards to make them lossless on a per-priority traffic basis, and to provide a mechanism to carry converged (LAN/SAN/IPC) traffic on a single physical link. Although CEE standards were designed with FCoE in mind, they are not limited to FCoE installations. CEE features can be utilized in traditional LAN (non-FCoE) networks to provide lossless guarantees on a per-priority basis, and to provide efficient bandwidth allocation based on application needs.

Turning CEE On or Off

By default on the CN4093, CEE is turned off. To turn CEE on or off, use the following ISCLI configuration mode commands:

```
CN 4093(config)# [no] cee enable
```

CAUTION:
Turning CEE on will automatically change some 802.1p QoS and 802.3x standard flow control settings on the CN4093. Read the following material carefully to determine whether you will need to take action to reconfigure expected settings.

It is recommended that you backup your configuration prior to turning CEE on. Viewing the file will allow you to manually re-create the equivalent configuration once CEE is turned on, and will also allow you to recover your prior configuration if you need to turn CEE off.

Effects on Link Layer Discovery Protocol

When CEE is turned on, Link Layer Discovery Protocol (LLDP) is automatically turned on and enabled for receiving and transmitting DCBX information. LLDP cannot be turned off while CEE is turned on.

Effects on 802.1p Quality of Service

While CEE is off (the default), the CN4093 allows 802.1p priority values to be used for Quality of Service (QoS) configuration (see "Quality of Service" on page 219). 802.1p QoS default settings are shown in Table 27 on page 307, but can be changed by the administrator.

When CEE is turned on, 802.1p QoS is replaced by ETS (see "Enhanced Transmission Selection" on page 320). As a result, while CEE is turned on, using 802.1p QoS configuration commands causes an error message to appear stating that the CEE global ETS command must be used to configure 802.1p-based QoS.

In addition, when CEE is turned on, prior 802.1p QoS settings are replaced with new defaults designed for use with ETS priority groups (PGIDs) as shown in Table 27:

Table 27. *CEE Effects on 802.1p Defaults*

802.1p QoS Configuration With CEE Off (default)		
Priority	**COSq**	**Weight**
0	0	1
1	1	2
2	2	3
3	3	4
4	4	5
5	5	7
6	6	15
7	7	0

ETS Configuration With CEE On		
Priority	**COSq**	**PGID**
0	0	0
1	0	0
2	0	0
3	1	1
4	2	2
5	2	2
6	2	2
7	2	2

When CEE is on, the default ETS configuration also allocates a portion of link bandwidth to each PGID as shown in Table 28:

Table 28. *Default ETS Bandwidth Allocation*

PGID	Typical Use	Bandwidth
0	LAN	10%
1	SAN	50%
2	Latency-sensitive LAN	40%

If the prior, non-CEE configuration used 802.1p priority values for different purposes, or does not expect bandwidth allocation as shown in Table 28 on page 307, when CEE is turned on, the administrator should reconfigure ETS settings as appropriate.

Each time CEE is turned on or off, the appropriate ETS or 802.1p QoS default settings shown in Table 27 on page 307 are restored, and any manual settings made to prior ETS or 802.1p QoS configurations are cleared.

It is recommended that a configuration backup be made prior to turning CEE on or off. Viewing the configuration file will allow the administrator to manually re-create the equivalent configuration under the new CEE mode, and will also allow for the recovery of the prior configuration if necessary.

Effects on Flow Control

When CEE is off (the default), 802.3x standard flow control is enabled on all switch ports by default.

When CEE is turned on, standard flow control is disabled on all ports, and in its place, PFC (see "Priority-Based Flow Control" on page 316) is enabled on all ports for 802.1p priority value 3. This default is chosen because priority value 3 is commonly used to identify FCoE traffic in a CEE environment and must be guaranteed lossless behavior. PFC is disabled for all other priority values.

It is recommend that a configuration backup be made prior to turning CEE on or off. Viewing the configuration file will allow the administrator to manually re-create the equivalent configuration under the new CEE mode, and will also allow for the recovery of the prior configuration if necessary.

When CEE is on, PFC can be enabled only on priority value 3 and one other priority. If flow control is required on additional priorities on any given port, consider using standard flow control on that port, so that regardless of which priority traffic becomes congested, a flow control frame is generated.

FCoE Initialization Protocol Snooping

FCoE Initialization Protocol (FIP) snooping is an FCoE feature. In order to enforce point-to-point links for FCoE traffic outside the regular Fibre Channel topology, Ethernet ports used in FCoE can be automatically and dynamically configured with Access Control Lists (ACLs).

Using FIP snooping, the CN4093 examines the FIP frames normally exchanged between the FCF and ENodes to determine information about connected FCoE devices. This information is used to create narrowly tailored ACLs that permit expected FCoE traffic to and from confirmed Fibre Channel nodes, and deny all other undesirable FCoE or FIP traffic.

FIP Snooping Requirements

The following are required for implementing the FIP snooping bridge feature:

- The CN4093 must be connected to the Fibre Channel network through a FCF such as a *Lenovo Rackswitch G8264CS*, another *Lenovo CN4093 10Gb Converged Scalable Switch* or a Cisco Nexus 5000 Series Switch.
- For each CN4093 switch port participating in FCoE, the connected server must use a FCoE-licensed Converged Network Adapter (CNA) and must have the FCoE license enabled (if applicable) on the CNA.
- CEE must be turned on (see "Turning CEE On or Off" on page 306). When CEE is on, the DCBX, PFC and ETS features are enabled and configured with default FCoE settings. These features may be reconfigured, but must remain enabled for FCoE to function.
- FIP snooping must be turned on (see "Global FIP Snooping Settings" on page 310). When FIP snooping is turned on, the feature is enabled on all ports by default. The administrator can disable FIP snooping on individual ports that do not require FCoE, but FIP snooping must remain enabled on all FCoE ports for FCoE to function.

Port Aggregation

Enterprise NOS 8.4 supports port aggregation for FCoE connections. Link Aggregation (LAGs) can be used for separate FCoE traffic, or for Ethernet and FCoE traffic. The ports can be grouped as static LAGS or dynamic LACP LAGs.

Normal aggregation operations such as creating or enabling a LAG, and adding or removing member ports can be performed. When a port is added to a LAG, FCFs previously detected on the port will be deleted. The deleted FCF may be relearned later. However, this may cause flickering in the network traffic. It is recommended that any LAG changes are made prior to live FCoE traffic.

Priority-based Flow Control (PFC) and Data Center Bridging (DCBX) are configured on a per-port basis. Each port in a LAG must have the same ETS, PFC and DCBX configuration. When a port ceases to be a LAG member, its configuration does not change.

Note: If the ports chosen to be part of a certain LAG do not have the same PFC, ETS or DCBX configurations, the switch will display an error.

Global FIP Snooping Settings

By default, the FIP snooping feature is turned off for the CN4093. The following commands are used to turn the feature on or off:

```
CN 4093(config)# [no] fcoe fips enable
```

Note: FIP snooping requires CEE to be turned on (see "Turning CEE On or Off" on page 306).

When FIP snooping is on, port participation may be configured on a port-by-port basis (see below).

When FIP snooping is off, all FCoE-related ACLs generated by the feature are removed from all switch ports.

FIP Snooping for Specific Ports

When FIP snooping is globally turned on, ports may be individually configured for participation in FIP snooping and automatic ACL generation. By default, FIP snooping is enabled for each port. To change the setting for any specific port, use the following CLI commands:

```
CN 4093(config)# [no] fcoe fips port <port number, alias, list, or range> enable
```

When FIP snooping is enabled on a port, FCoE-related ACLs will be automatically configured.

When FIP snooping is disabled on a port, all FCoE-related ACLs on the port are removed, and the switch will enforce no FCoE-related rules for traffic on the port.

Note: FIP Snooping and IPv6 ACLs are not support simultaneously on the same ports. To use FIP snooping, remove IPv6 ACLs from the port.

FIPS LAG Support on Server Ports

FIPS LAG Support allows FCoE and Ethernet traffic to co-exist within the same LAGs (ports). By default, FCoE servers (CNA/HBA) do not support aggregation, while Ethernet (NIC/CNA) are aggregation capable. Due to this incompatibility on FCoE capable servers, the FCoE traffic is generated on separate (exclusive) ports whenever Ethernet adapters need to be consolidated into a LAG.

FIPS LAG Support allows FCoE traffic and traditional Ethernet traffic to use the same ports for traffic by pinning each destination FCoE Enode-MAC to a static switch port within the LAG. This is due to each server port within a LAG expecting FCoE traffic with a destination - MAC as its Enode-MAC to arrive on the same port within the LAG from the switch (i.e. FCoE traffic with a destination Enode-MAC is always expected to traverse the link with that Enode-MAC). Initially, any incoming FIP packets are snooped by the switch and if the Enode-MAC is new (previously undiscovered) and source port is part of a LAG, then a static Enode-MAC entry is installed within the switch. Any unicast (FIP) response then onward is transmitted using the assigned port within the LAG and not any other port. Thus, FCoE traffic then strictly transmits across only the assigned port (within the LAG) for each Enode. Similarly, any VN-Port MACs are pinned on a port by port basis within a LAG. Regular (non-FCoE) Ethernet traffic will continue to operate across the LAG normally (using any of the links based on balancing algorithm).

This feature is automatically activated upon server-port LAG mode detection by FIPS.

Note: FCoE Fips must be enabled on FSB/FCF switch.

Port FCF and ENode Detection

When FIP snooping is enabled on a port, the port is placed in FCF auto-detect mode by default. In this mode, the port assumes connection to an ENode unless FIP packets show the port is connected to an external FCF.

Ports can also be specifically configured as to whether automatic FCF detection should be used, or whether the port is connected to an external FCF or ENode:

```
CN 4093(config)# fcoe fips port <port alias, number, list, or range> fcf-mode
{auto|on|off}
```

When FCF mode is `on`, the port is assumed to be connected to a trusted external FCF, and only ACLs appropriate to FCFs will be installed on the port. When `off`, the port is assumed to be connected to an ENode, and only ACLs appropriate to ENodes will be installed. When the mode is changed (either through manual configuration or as a result of automatic detection), the appropriate ACLs are automatically added, removed, or changed to reflect the new FCF or ENode connection.

FCoE Connection Timeout

FCoE-related ACLs and VLANs are added, changed, and removed as FCoE device connection and disconnection are discovered. In addition, the administrator can enable or disable automatic removal of ACLs and VLANs for FCFs and other FCoE connections that timeout (fail or are disconnected) without FIP notification.

By default, automatic removal of ACLs upon timeout is enabled. To change this function, use the following CLI command:

```
CN 4093(config)# [no] fcoe fips timeout-acl
```

FCoE ACL Rules

When FIP Snooping is enabled on a port, the switch automatically installs the appropriate ACLs to enforce the following rules for FCoE traffic:

- Ensure that FIP frames from ENodes may only be addressed to FCFs.
- Flag important FIP packets for switch processing.
- Ensure no end device uses an FCF MAC address as its source.
- Each FCoE port is assumed to be connected to an ENode and include ENode-specific ACLs installed, until the port is either detected or configured to be connected to an external FCF.
- Ports that are configured to have FIP snooping disabled will not have any FIP or FCoE related ACLs installed.
- Prevent transmission of all FCoE frames from an ENode prior to its successful completion of login (FLOGI) to the FCF.
- After successful completion of FLOGI, ensure that the ENode uses only those FCoE source addresses assigned to it by the FCF.
- After successful completion of FLOGI, ensure that all ENode FCoE source addresses originate from or are destined to the appropriate ENode port.
- After successful completion of each FLOGI, ensure that FCoE frames may only be addressed to the FCFs that accept them.

Initially, a basic set of FCoE-related ACLs will be installed on all ports where FIP snooping is enabled. As the switch encounters FIP frames and learns about FCFs and ENodes that are attached or disconnect, ACLs are dynamically installed or expanded to provide appropriate security.

When an FCoE connection logs out, or times out (if ACL timeout is enabled), the related ACLs will be automatically removed.

FCoE-related ACLs are independent of manually configured ACLs used for regular Ethernet purposes. FCoE ACLs generally have a higher priority over standard ACLs.

FCoE VLANs

Before the switch applies FIP Snooping, all internal switch ports connected to ENodes and all external ports connected to external FCFs should be members of at least one common VLAN (for example, VLAN 1). This allows the ENode CNA and the external FCF to exchange initial FIP VLAN request and notification packets. Once FIP Snooping is applied, FCoE packets are exchanged using one configured FCoE VLAN for each attached FCF.

Each ENode port must have VLAN tagging enabled, and must belong to the same VLAN as the external FCF to which it will connect. In topologies where a single external FCF is connected to the switch, all ENode and FCF ports belong to the same VLAN (typically VLAN 1002). When multiple FCFs are connected to the switch, each FCF must be assigned a unique VLAN, and each ENode must be assigned to the VLAN for only one particular FCF.

The administrator must ensure that the VLAN configured for each FCF and its ENodes is supported by the participating FCF and ENode CNAs.

Note: The FCoE MAC Address Prefix (FC-Map) value is used to identify traffic on a FCoE VLAN. The valid FC-Map values are from 0xefcf00 to 0x0efcff and are configured automatically for each FCoE vlan. If you need to manually configure the FC-MAP, use values in the range 0xefcf00 to 0x0efcf4. The other FC-Map values are reserved.

Viewing FIP Snooping Information

ACLs automatically generated under FIP snooping are independent of regular, manually configure ACLs, and are not listed with regular ACLs in switch information and statistics output. Instead, FCoE ACLs are shown using the following CLI commands:

```
CN 4093# show fcoe fips information                (Show all FIP-related information)
CN 4093# show fcoe fips port <ports> information   (Show FIP info for a selected port)
```

For example:

```
CN 4093# show fcoe fips port ext4 information

FIP Snooping on port ext4:
This port has been detected to be an FCF port.

FIPS ACLs configured on this port:
Ethertype 0x8914, action permit.
dmac 00:00:18:01:00:XX, Ethertype 0x8914, action permit.
```

For each ACL, the required traffic criteria are listed, along with the action taken (permit or deny) for matching traffic. ACLs are listed in order of precedence and evaluated in the order shown.

The administrator can also view other FCoE information:

```
CN 4093# show fcoe fips fcf          (Show all detected FCFs)
CN 4093# show fcoe fips fcoe         (Show all FCoE connections)
```

Operational Commands

The administrator may use the operational commands to delete FIP-related entries from the switch.

To delete a specific FCF entry and all associated ACLs from the switch, use the following command:

```
CN 4093# no fcoe fips fcf <FCF MAC address> [<VLAN number>]
```

FIP Snooping Configuration

In this example, as shown in Figure 32 on page 304, port INTA1 is connected to an ENode, and EXT22 is connected to the Fibre Channel network via an internal FCF (see"Fibre Channel" on page 333). FIP snooping can be configured on these ports using the following CLI commands:

1. Enable VLAN tagging on FCoE ports:

```
CN 4093(config)# interface port INTA1, EXT22    (Select FCoE port)
CN 4093(config-if)# switchport mode trunk       (Enable VLAN tagging)
CN 4093(config-if)# exit                         (Exit port configuration mode)
```

2. Place FCoE ports into a VLAN supported by the FCF and CNAs (typically VLAN 1002):

```
CN 4093(config)# vlan 1002
CN 4093(config-vlan)# exit
CN 4093(config)# interface port INTA1, EXT22
CN 4093(config-if)# switchport trunk allowed vlan 1002
CN 4093(config-if)# exit
```

Note: Placing ports into the VLAN *after* tagging is enabled helps to ensure that their port VLAN ID (PVID) is not accidentally changed.

3. Turn CEE on.

```
CN 4093(config)# cee enable
```

Note: Turning CEE on will automatically change some 802.1p QoS and 802.3x standard flow control settings and menus (see "Turning CEE On or Off" on page 306).

4. Turn global FIP snooping on:

```
CN 4093(config)# fcoe fips enable
```

5. Disable FIP snooping on all non-FCoE external ports:

```
CN 4093(config)# no fcoe fips port INTA2-EXT21 enable
```

6. Enable FIP snooping on FCoE ports, and set the desired FCF mode:

```
CN 4093(config)# fcoe fips port INTA1 enable          (Enable FIPS on FCoE ports)
CN 4093(config)# fcoe fips port INTA1 fcf-mode off    (Set as ENode connection)
CN 4093(config)# fcoe fips port EXT22 fcf-mode on     (Set as FCF connection)
```

Note: By default, FIP snooping is enabled on all ports and the FCF mode set for automatic detection. The configuration in this step is unnecessary if default settings have not been changed, and is shown merely as a manual configuration example.

7. Save the configuration.

Priority-Based Flow Control

Priority-based Flow Control (PFC) is defined in IEEE 802.1Qbb. PFC extends the IEEE 802.3x standard flow control mechanism. Under standard flow control, when a port becomes busy, the switch manages congestion by pausing all the traffic on the port, regardless of the traffic type. PFC provides more granular flow control, allowing the switch to pause specified types of traffic on the port, while other traffic on the port continues.

PFC pauses traffic based on 802.1p priority values in the VLAN tag. The administrator can assign different priority values to different types of traffic and then enable PFC for up to two specific priority values: priority value 3, and one other. The configuration can be applied on a port-by-port basis, or globally for all ports on the switch. Then, when traffic congestion occurs on a port (caused when ingress traffic exceeds internal buffer thresholds), only traffic with priority values where PFC is enabled is paused. Traffic with priority values where PFC is disabled proceeds without interruption but may be subject to loss if port ingress buffers become full.

Although PFC is useful for a variety of applications, it is required for FCoE implementation where storage (SAN) and networking (LAN) traffic are converged on the same Ethernet links. Typical LAN traffic tolerates Ethernet packet loss that can occur from congestion or other factors, but SAN traffic must be lossless and requires flow control.

For FCoE, standard flow control would pause both SAN and LAN traffic during congestion. While this approach would limit SAN traffic loss, it could degrade the performance of some LAN applications that expect to handle congestion by dropping traffic. PFC resolves these FCoE flow control issues. Different types of SAN and LAN traffic can be assigned different IEEE 802.1p priority values. PFC can then be enabled for priority values that represent SAN and LAN traffic that must be paused during congestion, and disabled for priority values that represent LAN traffic that is more loss-tolerant.

PFC requires CEE to be turned on ("Turning CEE On or Off" on page 306). When CEE is turned on, PFC is enabled on priority value 3 by default. Optionally, the administrator can also enable PFC on one other priority value, providing lossless handling for another traffic type, such as for a business-critical LAN application.

Note: For any given port, only one flow control method can be implemented at any given time: either PFC or standard IEEE 802.3x flow control.

Global vs. Port-by-Port PFC Configuration

PFC requires CEE to be turned on ("Turning CEE On or Off" on page 306). When CEE is turned on, standard flow control is disabled on all ports, and PFC is enabled on all ports for 802.1p priority value 3. While CEE is turned on, PFC cannot be disabled for priority value 3. This default is chosen because priority value 3 is commonly used to identify FCoE traffic in a CEE environment and must be guaranteed lossless behavior. PFC is disabled for all other priority values by default, but can be enabled for one additional priority value.

The administrator can also configure PFC on a port-by-port basis. The method used will typically depend on the following:

- Port-by-port PFC configuration is desirable in most mixed environments where some CN4093 ports are connected to CEE-capable (FCoE) switches, gateways, and Converged Network Adapters (CNAs), and other CN4093 ports are connected to non-CEE Layer 2/Layer 3 switches, routers and Network Interface Cards (NICs).
- Global PFC configuration is preferable in networks that implement end-to-end CEE devices. For example, if all ports are involved with FCoE and can use the same SAN and LAN priority value configuration with the same PFC settings, global configuration is easy and efficient.
- Global PFC configuration can also be used in some mixed environments where traffic with PFC-enabled priority values occurs only on ports connected to CEE devices, and not on any ports connected to non-CEE devices. In such cases, PFC can be configured globally on specific priority values even though not all ports make use them.
- PFC is not restricted to CEE and FCoE networks. In any LAN where traffic is separated into different priorities, PFC can be enabled on priority values for loss-sensitive traffic. If all ports have the same priority definitions and utilize the same PFC strategy, PFC can be globally configured.

If you want to enable PFC on a priority, do one of the following:

- Create a separate priority group (see "Priority Groups" on page 321).
- Move the priority to an existing priority group in which PFC is turned on.

  Since there are separate COS queue and ETS configurations, creating a distinct priority group is preferred.

When configuring ETS and PFC on the switch, perform ETS configuration before performing PFC configuration.

If two priorities are enabled on a port, the switch sends PFC frames for both priorities, even if only traffic tagged with one of the priorities is being received on that port.

Note: When using global PFC configuration in conjunction with the ETS feature (see "Enhanced Transmission Selection" on page 320), ensure that only pause-tolerant traffic (such as lossless FCoE traffic) is assigned priority values where PFC is enabled. Pausing other types of traffic can have adverse effects on LAN applications that expect uninterrupted traffic flow and tolerate dropping packets during congestion. Use PFC globally only if all priority values assigned for lossless traffic on one or more ports does not carry loss-tolerant traffic on other ports.

PFC Configuration Example

Note: DCBX may be configured to permit sharing or learning PFC configuration with or from external devices. This example assumes that PFC configuration is being performed manually. See "Data Center Bridging Capability Exchange" on page 326 for more information on DCBX.

This example is consistent with the network shown in Figure 32 on page 304. In this example, the following topology is used.

Table 29. *Port-Based PFC Configuration*

Switch Port	802.1p Priority	Usage	PFC Setting
EXT1	0-2	LAN	Disabled
	3	(not used)	**Enabled**
	4	Business-critical LAN	**Enabled**
	others	(not used)	Disabled
EXT22	3	Fiber Channel network	**Enabled**
INTA1	3	FCoE	**Enabled**
	others	(not used)	Disabled
INTA2	0-2	LAN	Disabled
	3	(not used)	**Enabled**
	4	Business-critical LAN	**Enabled**
	others	(not used)	Disabled

In this example, PFC is to facilitate lossless traffic handling for FCoE (priority value 3) and a business-critical LAN application (priority value 4).

Assuming that CEE is off (the CN4093 default), the example topology shown in Table 29 on page 318 can be configured using the following commands:

1. Turn CEE on.

```
CN 4093(config)# cee enable
```

Note: Turning CEE on will automatically change some 802.1p QoS and 802.3x standard flow control settings and menus (see "Turning CEE On or Off" on page 306).

2. Enable PFC for the FCoE traffic.

Note: PFC is enabled on priority 3 by default. If using the defaults, the manual configuration commands shown in this step are not necessary.

```
CN 4093(config)# cee port INTA1 pfc priority 3 enable   (FCoE priority)
CN 4093(config)# cee port INTA1 pfc priority 3 description "FCoE" (Optional)
CN 4093(config)# cee port EXT22 pfc priority 3 enable   (FCoE priority)
CN 4093(config)# cee port EXT22 pfc priority 3 description "FCoE" (Optional)
```

3. Enable PFC for the business-critical LAN application:

```
CN 4093(config)# cee port INTA2 pfc priority 4 enable(LAN priority)
CN 4093(config)# cee port INTA2 pfc priority 4 description "Critical LAN"
CN 4093(config)# cee port EXT1 pfc priority 4 enable( LAN priority)
CN 4093(config)# cee port EXT1 pfc priority 4 description "Critical LAN"
```

4. Save the configuration.

Enhanced Transmission Selection

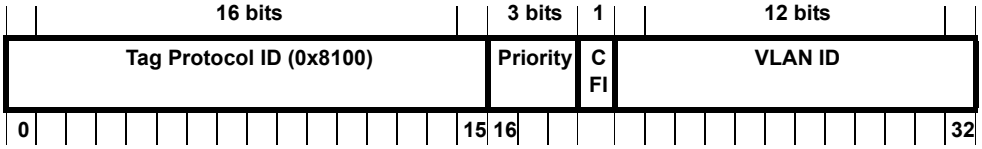
Enhanced Transmission Selection (ETS) is defined in IEEE 802.1Qaz. ETS provides a method for allocating port bandwidth based on 802.1p priority values in the VLAN tag. Using ETS, different amounts of link bandwidth can specified for different traffic types (such as for LAN, SAN, and management).

ETS is an essential component in a CEE environment that carries different types of traffic, each of which is sensitive to different handling criteria, such as Storage Area Networks (SANs) that are sensitive to packet loss, and LAN applications that may be latency-sensitive. In a single converged link, such as when implementing FCoE, ETS allows SAN and LAN traffic to coexist without imposing contrary handling requirements upon each other.

The ETS feature requires CEE to be turned on (see "Turning CEE On or Off" on page 306).

802.1p Priority Values

Under the 802.1p standard, there are eight available priority values, with values numbered 0 through 7, which can be placed in the priority field of the 802.1Q VLAN tag:



Servers and other network devices may be configured to assign different priority values to packets belonging to different traffic types (such as SAN and LAN).

ETS uses the assigned 802.1p priority values to identify different traffic types. The various priority values are assigned to priority groups (PGID), and each priority group is assigned a portion of available link bandwidth.

Priority values within in any specific ETS priority group are expected to have similar traffic handling requirements with respect to latency and loss.

802.1p priority values may be assigned by the administrator for a variety of purposes. However, when CEE is turned on, the CN4093 sets the initial default values for ETS configuration as follows:

Figure 33. Default ETS Priority Groups

Typical Traffic Type	802.1p Priority	PGID	Bandwidth Allocation
LAN	0	0	10%
LAN	1		
LAN	2		
SAN	3	1	50%
Latency-Sensitive LAN	4	2	40%
Latency-Sensitive LAN	5		
Latency-Sensitive LAN	6		
Latency-Sensitive LAN	7		

In the assignment model shown in Figure 33 on page 320, priorities values 0 through 2 are assigned for regular Ethernet traffic, which has "best effort" transport characteristics.

Priority 3 is typically used to identify FCoE (SAN) traffic.

Priorities 4-7 are typically used for latency sensitive traffic and other important business applications. For example, priority 4 and 5 are often used for video and voice applications such as IPTV, Video on Demand (VoD), and Voice over IP (VoIP). Priority 6 and 7 are often used for traffic characterized with a "must get there" requirement, with priority 7 used for network control which is requires guaranteed delivery to support configuration and maintenance of the network infrastructure.

Note: The default assignment of 802.1p priority values on the CN4093 changes depending on whether CEE is on or off. See "Turning CEE On or Off" on page 306 for details.

Priority Groups

For ETS use, each 801.2p priority value is assigned to a priority group which can then be allocated a specific portion of available link bandwidth. To configure a priority group, the following is required:

- CEE must be turned on ("Turning CEE On or Off" on page 306) for the ETS feature to function.
- A priority group must be assigned a priority group ID (PGID), one or more 802.1p priority values, and allocated link bandwidth greater than 0%.

PGID

Each priority group is identified with number (0 through 7, and 15) known as the PGID.

PGID 0 through 7 may each be assigned a portion of the switch's available bandwidth.

PGID 8 through 14 are reserved as per the 802.1Qaz ETS standard.

PGID 15 is a strict priority group. It is generally used for critical traffic, such as network management. Any traffic with priority values assigned to PGID 15 is permitted as much bandwidth as required, up to the maximum available on the switch. After serving PGID 15, any remaining link bandwidth is shared among the other groups, divided according to the configured bandwidth allocation settings.

All 802.1p priority values assigned to a particular PGID should have similar traffic handling requirements. For example, PFC-enabled traffic should not be grouped with non-PFC traffic. Also, traffic of the same general type should be assigned to the same PGID. Splitting one type of traffic into multiple 802.1p priorities, and then assigning those priorities to different PGIDs may result in unexpected network behavior.

Each 802.1p priority value may be assigned to only one PGID. However, each PGID may include multiple priority values. Up to eight PGIDs may be configured at any given time.

Assigning Priority Values to a Priority Group

Each priority group may be configured from its corresponding ETS Priority Group, available using the following command:

```
CN 4093(config)# cee global ets priority-group pgid <group number (0-7, or 15)>
priority <priority list>
```

where *priority list* is one or more 802.1p priority values (with each separated by a comma). For example, to assign priority values 0 through 2:

```
CN 4093(config)# cee global ets priority-group pgid <group number (0-7, or 15)>
priority 0,1,2
```

Note: Within any specific PGID, the PFC settings (see "Priority-Based Flow Control" on page 316) should be the same (enabled or disabled) for all priority values within the group. PFC can be enabled only on priority value 3 and one other priority. If the PFC setting is inconsistent within a PGID, a warning message is reported when attempting to apply the configuration.

When assigning priority values to a PGID, the specified priority value will be automatically removed from its old group and assigned to the new group when the configuration is applied.

Each priority value must be assigned to a PGID. Priority values may not be deleted or unassigned. To remove a priority value from a PGID, it must be moved to another PGID.

For PGIDs 0 through 7, bandwidth allocation can also be configured through the ETS Priority Group menu. See for "Allocating Bandwidth" on page 323 for details.

Note: In a stacking setup, when there are multiple priorities assigned to the same low-bandwidth PG and the PG traffic is composed of various packet sizes;, a marginal offset from the configured PG bandwidth may be observed. However, the final bandwidth ratios among all PGs observed at the destination port is still correct according to the ETS configuration.

Deleting a Priority Group

A priority group is automatically deleted when it contains no associated priority values, and its bandwidth allocation is set to 0%.

Note: The total bandwidth allocated to PGID 0 through 7 must equal exactly 100%. Reducing the bandwidth allocation of any group will require increasing the allocation to one or more of the other groups (see "Allocating Bandwidth" on page 323).

Allocating Bandwidth

Allocated Bandwidth for PGID 0 Through 7

The administrator may allocate a portion of the switch's available bandwidth to PGIDs 0 through 7. Available bandwidth is defined as the amount of link bandwidth that remains after priorities within PGID 15 are serviced (see "Unlimited Bandwidth for PGID 15" on page 323), and assuming that all PGIDs are fully subscribed. If any PGID does not fully consume its allocated bandwidth, the unused portion is made available to the other priority groups.

Priority group bandwidth allocation can be configured using the following command:

```
CN 4093(config)# cee global ets priority-group pgid <priority group number>
bandwidth <bandwidth allocation> pgid <priority group number> bandwidth
<bandwidth allocation>
```

where *bandwidth allocation* represents the percentage of link bandwidth, specified as a number between 0 and 100, in 1% increments.

The following bandwidth allocation rules apply:

- Bandwidth allocation must be 0% for any PGID that has no assigned 802.1p priority values.
- Any PGID assigned one or more priority values must have a bandwidth allocation greater than 0%.
- Total bandwidth allocation for groups 0 through 7 must equal exactly 100%. Increasing or reducing the bandwidth allocation of any PGID also requires adjusting the allocation of other PGIDs to compensate.

If these conditions are not met, the switch will report an error when applying the configuration.

Note: Actual bandwidth used by any specific PGID may vary from configured values by up to 10% of the available bandwidth in accordance with 802.1Qaz ETS standard. For example, a setting of 10% may be served anywhere from 0% to 20% of the available bandwidth at any given time.

Unlimited Bandwidth for PGID 15

PGID 15 is permitted unlimited bandwidth and is generally intended for critical traffic (such as switch management). Traffic in this group is given highest priority and is served before the traffic in any other priority group.

If PGID 15 has low traffic levels, most of the switch's bandwidth will be available to serve priority groups 0 through 7. However, if PGID 15 consumes a larger part of the switch's total bandwidth, the amount available to the other groups is reduced.

Note: Consider traffic load when assigning priority values to PGID 15. Heavy traffic in this group may restrict the bandwidth available to other groups.

Configuring ETS

Consider an example consistent with that used for port-based PFC configuration (on page 318):

Table 30. *ETS Configuration*

Priority	Usage	PGID	Bandwidth
0	LAN (best effort delivery)	0	10%
1	LAN (best effort delivery)		
2	LAN (best effort delivery)		
3	SAN (Fibre Channel over Ethernet, with PFC)	1	20%
4	Business Critical LAN (lossless Ethernet, with PFC)	2	30%
5	Latency-sensitive LAN	3	40%
6	Latency-sensitive LAN		
7	Network Management (strict)	15	unlimited

The example shown in Table 30 is only slightly different than the default configuration shown in Figure 33 on page 320. In this example, latency-sensitive LAN traffic (802.1p priority 5 through 6) are moved from priority group 2 to priority group 3. This leaves Business Critical LAN traffic (802.1p priority 4) in priority group 2 by itself. Also, a new group for network management traffic has been assigned. Finally, the bandwidth allocation for priority groups 1, 2, and 3 are revised.

Note: DCBX may be configured to permit sharing or learning PFC configuration with or from external devices. This example assumes that PFC configuration is being performed manually. See "Data Center Bridging Capability Exchange" on page 326 for more information on DCBX.

This example can be configured using the following commands:

1. Turn CEE on.

```
CN 4093(config)# cee enable
```

Note: Turning CEE on will automatically change some 802.1p QoS and 802.3x standard flow control settings (see "Turning CEE On or Off" on page 306).

2. Configure each allocated priority group with a description (optional), list of 802.1p priority values, and bandwidth allocation:

```
CN 4093(config)# cee global ets priority-group pgid 0 priority 0,1,2
(Select a group for regular LAN, and  set for 802.1p priorities 0, 1, and 2)

CN 4093(config)# cee global ets priority-group pgid 0 description
"Regular LAN"
(Set a group description—optional)

CN 4093(config)# cee global ets priority-group pgid 1 priority 3
(Select a group for SAN traffic, and set for 802.1p priority 3)

CN 4093(config)# cee global ets priority-group pgid 1 description "SAN"
(Set a group description—optional)

CN 4093(config)# cee global ets priority-group pgid 2 priority 4
(Select a group for latency traffic and set for 802.1p priority 4)

CN 4093(config)# cee global ets priority-group pgid 2 description
"Biz-Critical LAN"
(Set a group description—optional)

CN 4093(config)# cee global ets priority-group pgid 3 description
"Latency-Sensitive LAN"
(Set a group description—optional)

CN 4093(config)# cee global ets priority-group pgid 3 priority 5,6 pgid 0
bandwidth 10 pgid 1 bandwidth 20 pgid 2 bandwidth 30 pgid 3 bandwidth 40
(Configure link bandwidth restriction)
```

3. Configure the strict priority group with a description (optional) and a list of 802.1p priority values:

```
CN 4093(config)# cee global ets priority-group pgid 15 priority 7
CN 4093(config)# cee global ets priority-group pgid 15 description
"Network Management"
```

Note: Priority group 15 is permitted unlimited bandwidth. As such, the commands for priority group 15 do not include bandwidth allocation.

4. Save the configuration.

Data Center Bridging Capability Exchange

Data Center Bridging Capability Exchange (DCBX) protocol is a vital element of CEE. DCBX allows peer CEE devices to exchange information about their advanced capabilities. Using DCBX, neighboring network devices discover their peers, negotiate peer configurations, and detect misconfigurations.

DCBX provides two main functions on the CN4093:

- Peer information exchange

  The switch uses DCBX to exchange information with connected CEE devices. For normal operation of any FCoE implementation on the CN4093, DCBX must remain enabled on all ports participating in FCoE.

- Peer configuration negotiation

  DCBX also allows CEE devices to negotiate with each other for the purpose of automatically configuring advanced CEE features such as PFC, ETS, and (for some CNAs) FIP. The administrator can determine which CEE feature settings on the switch are communicated to and matched by CEE neighbors, and also which CEE feature settings on the switch may be configured by neighbor requirements.

The DCBX feature requires CEE to be turned on (see "Turning CEE On or Off" on page 306).

DCBX Settings

When CEE is turned on, DCBX is enabled for peer information exchange on all ports. For configuration negotiation, the following default settings are configured:

- Application Protocol: FCoE and FIP snooping is set for traffic with 802.1p priority 3
- PFC: Enabled on 802.1p priority 3
- ETS
  - Priority group 0 includes priority values 0 through 2, with bandwidth allocation of 10%
  - Priority group 1 includes priority value 3, with bandwidth allocation of 50%
  - Priority group 2 includes priority values 4 through 7, with bandwidth allocation of 40%

Enabling and Disabling DCBX

When CEE is turned on, DCBX can be enabled and disabled on a per-port basis, using the following commands:

```
CN 4093(config)# [no] cee port <port alias or number> dcbx enable
```

When DCBX is enabled on a port, Link Layer Detection Protocol (LLDP) is used to exchange DCBX parameters between CEE peers. Also, the interval for LLDP transmission time is set to one second for the first five initial LLDP transmissions, after which it is returned to the administratively configured value. The minimum delay between consecutive LLDP frames is also set to one second as a DCBX default.

Peer Configuration Negotiation

CEE peer configuration negotiation can be set on a per-port basis for a number of CEE features. For each supported feature, the administrator can configure two independent flags:

- The `advertise` flag

  When this flag is set for a particular feature, the switch settings will be transmit to the remote CEE peer. If the peer is capable of the feature, and willing to accept the CN4093 settings, it will be automatically reconfigured to match the switch.

- The `willing` flag

  Set this flag when required by the remote CEE peer for a particular feature as part of DCBX signaling and support. Although some devices may also expect this flag to indicate that the switch will accept overrides on feature settings, the CN4093 retains its configured settings. As a result, the administrator should configure the feature settings on the switch to match those expected by the remote CEE peer.

These flags are available for the following CEE features:

- Application Protocol

  DCBX exchanges information regarding FCoE and FIP snooping, including the 802.1p priority value used for FCoE traffic. The `advertise` flag is set or reset using the following command:

  ```
  CN 4093(config)# [no] cee port <port alias or number> dcbx app_proto
  advertise
  ```

  The `willing` flag is set or reset using the following command:

  ```
  CN 4093(config)# [no] cee port <port alias or number> dcbx app_proto willing
  ```

- PFC

  DCBX exchanges information regarding whether PFC is enabled or disabled on the port. The `advertise` flag is set or reset using the following command:

```
CN 4093(config)# [no] cee port <port alias or number> dcbx pfc advertise
```

  The willing flag is set or reset using the following command:

```
CN 4093(config)# [no] cee port <port alias or number> dcbx pfc willing
```

- ETS

  DCBX exchanges information regarding ETS priority groups, including their 802.1p priority members and bandwidth allocation percentages. The `advertise` flag is set or reset using the following command:

```
CN 4093(config)# [no] cee port <port alias or number> dcbx ets advertise
```

  The `willing` flag is set or reset using the following command:

```
CN 4093(config)# [no] cee port <port alias or number> dcbx ets willing
```

Configuring DCBX

Consider an example consistent Figure 32 on page 304 and used with the previous FCoE examples in this chapter:

- FCoE is used on port INTA1.
- CEE features are also used with LANs on ports INTA2 and EXT1.
- All other ports are disabled or are connected to non-CEE devices.

In this example, the CN4093 acts as the central point for CEE configuration. FCoE-related ports will be configured for advertising CEE capabilities, but not to accept external configuration. Other LAN ports that use CEE features will also be configured to advertise feature settings to remote peers, but not to accept external configuration. DCBX will be disabled on all non-CEE ports.

This example can be configured using the following commands:

1. Turn CEE on.

```
CN 4093(config)# cee enable
```

   Note: Turning CEE on will automatically change some 802.1p QoS and 802.3x standard flow control settings and menus (see "Turning CEE On or Off" on page 306).

2. Enable desired DCBX configuration negotiation on FCoE ports:

```
CN 4093(config)# cee port INTA1 dcbx enable
CN 4093(config)# cee port INTA1 dcbx app_proto advertise
CN 4093(config)# cee port INTA1 dcbx ets advertise
CN 4093(config)# cee port INTA1 dcbx pfc advertise
```

3. Enable desired DCBX advertisements on other CEE ports:

```
CN 4093(config)# cee port INTA2 dcbx enable
CN 4093(config)# cee port INTA2 dcbx app_proto advertise
CN 4093(config)# cee port INTA2 dcbx ets advertise
CN 4093(config)# cee port INTA2 dcbx pfc advertise

CN 4093(config)# cee port EXT1 dcbx enable
CN 4093(config)# cee port EXT1 dcbx app_proto advertise
CN 4093(config)# cee port EXT1 dcbx ets advertise
CN 4093(config)# cee port EXT1 dcbx pfc advertise
```

4. Disable DCBX for each non-CEE port as appropriate:

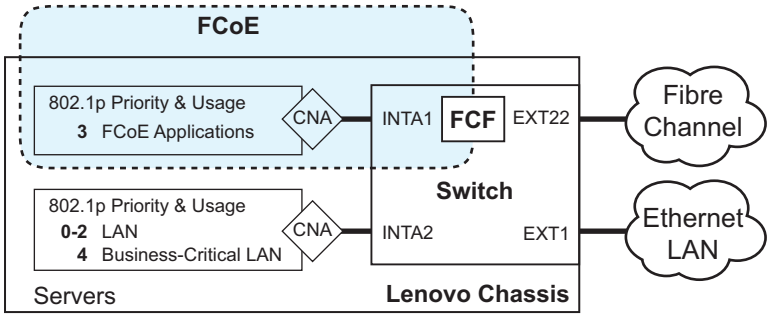
```
CN 4093(config)# no cee port INTA3-INTC14,EXT2-EXT22 dcbx enable
```

5. Save the configuration.

FCoE Example Configuration

The following example collects the various components from previous sections of this chapter.

Figure 34. A Mixed Fibre Channel and FCoE Network



In Figure 34 on page 330, a Fibre Channel network is connected to the CN4093 on port EXT22. The FCoE-enabled CN4093 is internally connected to a blade server (ENode) through an FCoE-enabled CNA on port INTA1. An internal FCF bridges the networks.

1. Configure FCoE ports as trunk ports and add them to FCoE VLAN for FCoE traffic and any Native VLAN (other than the FCoE VLAN) for FIP negotiation:

```
CN 4093(config)# interface port INTA1                (Select FCoE ports)
CN 4093(config-if)# switchport mode trunk            (Enable VLAN tagging)
CN 4093(config-if)# switchport trunk allowed vlan 1,1002 (Add VLANs)
CN 4093(config-if)# exit                             (Exit port configuration mode)
```

2. Turn CEE on.

```
CN 4093(config)# cee enable
```

Note: Turning CEE on will automatically change some 802.1p QoS and 802.3x standard flow control settings and menus (see "Turning CEE On or Off" on page 306).

3. Turn global FIP snooping on:

```
CN 4093(config)# fcoe fips enable
```

4. Disable FIP snooping on all non-FCoE external ports:

```
CN 4093(config)# no fcoe fips port EXT1-EXT20 enable
```

5. Enable FIP snooping on FCoE ports, and set the desired FCF mode:

```
CN 4093(config)# fcoe fips port INTA1 enable
CN 4093(config)# fcoe fips port INTA1 fcf-mode off   (Set as ENode connection)
```

Note: By default, FIP snooping is enabled on all ports and the FCF mode set for automatic detection. The configuration in this step is unnecessary if default settings have not been changed, and is shown merely as a manual configuration example.

6. Enable PFC for the FCoE traffic.

   Note: PFC is enabled on priority 3 by default. If using the defaults, the manual configuration commands shown in this step are not necessary.

```
CN 4093(config)# cee port INTA1 pfc priority 3 enable    (FCoE priority)
CN 4093(config)# cee port INTA1 pfc priority 3 description "FCoE" (Optional)
```

7. Enable PFC for the business-critical LAN application:

```
CN 4093(config)# cee port INTA2 pfc priority 4 enable     (LAN priority)
CN 4093(config)# cee port INTA2 pfc priority 4 description "Critical LAN"
CN 4093(config)# cee port EXT1 pfc priority 4 enable      (LAN priority)
CN 4093(config)# cee port EXT1 pfc priority 4 description "Critical LAN"
```

8. Configure each allocated priority group with a description (optional), list of 802.1p priority values, and bandwidth allocation:

```
CN 4093(config)# cee global ets priority-group pgid 0 priority 0,1,2
(Select a group for regular LAN, and set for 802.1p priorities 0, 1, and 2)

CN 4093(config)# cee global ets priority-group pgid 0 description
"Regular LAN"
(Set a group description—optional)

CN 4093(config)# cee global ets priority-group pgid 1 priority 3
(Select a group for SAN traffic, and set for 802.1p priority 3)

CN 4093(config)# cee global ets priority-group pgid 1 description "SAN"
(Set a group description—optional)

CN 4093(config)# cee global ets priority-group pgid 2 priority 4
(Select a group for latency traffic and set for 802.1p priority 4)

CN 4093(config)# cee global ets priority-group pgid 2 description
"Critical LAN"
(Set a group description—optional)

CN 4093(config)# cee global ets priority-group pgid 3 description
"Latency-Sensitive LAN"
(Set a group description—optional)

CN 4093(config)# cee global ets priority-group pgid 3 priority 5,6 pgid 0
bandwidth 10 pgid 1 bandwidth 20 pgid 2 bandwidth 30 pgid 3 bandwidth 40
(Configure link bandwidth restriction)
```

9. Configure the strict priority group with a description (optional) and a list of 802.1p priority values:

```
CN 4093(config)# cee global ets priority-group pgid 15 priority 7
CN 4093(config)# cee global ets priority-group pgid 15 description
"Network Management"
```

   Note: Priority group 15 is permitted unlimited bandwidth. As such, the commands for priority group 15 do not include bandwidth allocation.

10. Enable desired DCBX configuration negotiation on FCoE ports:

```
CN 4093(config)# cee port INTA1 dcbx enable
CN 4093(config)# cee port INTA1 dcbx app_proto advertise
CN 4093(config)# cee port INTA1 dcbx ets advertise
CN 4093(config)# cee port INTA1 dcbx pfc advertise
```

11. Enable desired DCBX advertisements on other CEE ports:

```
CN 4093(config)# cee port INTA2 dcbx enable
CN 4093(config)# cee port INTA2 dcbx app_proto advertise
CN 4093(config)# cee port INTA2 dcbx ets advertise
CN 4093(config)# cee port INTA2 dcbx pfc advertise

CN 4093(config)# cee port EXT1 dcbx enable
CN 4093(config)# cee port EXT1 dcbx app_proto advertise
CN 4093(config)# cee port EXT1 dcbx ets advertise
CN 4093(config)# cee port EXT1 dcbx pfc advertise
```

12. Disable DCBX for each non-CEE port as appropriate:

```
CN 4093(config)# no cee port INTA3-INTC14,EXT2-EXT22 dcbx enable
```

13. Configure the Fibre Channel network:

```
CN 4093(config)# system port ext21-ext22 type fc
CN 4093(config-vlan)# interface fc EXT22
CN 4093(config-if)# switchport trunk allowed vlan 1,1002
CN 4093(config-if)# exit

CN 4093(config)# vlan 1002
CN 4093(config-vlan)# npv enable
CN 4093(config-vlan)# npv traffic-map external-interface EXT22
CN 4093(config-vlan)# exit
```

Note: The Fibre Channel network is configured as an NPV gateway as described in "Fibre Channel" on page 333. Although VLAN properties for Fibre Channel and FCoE can be configured together, the additional Fibre Channel elements for this configuraiton are included at the end of this example in order to focus on the FCoE steps.

14. Save the configuration.

Chapter 18. Fibre Channel

This chapter describes how to configure the CN4093 for use with Fibre Channel networks.

Ethernet vs. Fibre Channel

As a converged switch, the CN4093 10 Gb Converged Scalable Switch provides simultaneous support of Ethernet and Fibre Channel networks.

Ethernet is ubiquitous in modern networks. It is generally quick, easy, and inexpensive to implement. Ethernet is also flexible and dynamic by nature. Devices join and leave a well-designed Ethernet network with little impact beyond their individual function. Because flux is the norm, Ethernet is classified as a "best effort" delivery protocol. This means that some loss of packets is acceptable, and that with multiple routes often available, packets in a stream may arrive at their destination out of sequence. Ethernet devices are expected to re-request and resend lost packets, and reassemble data in the proper order at the destination.

The Fibre Channel protocol adheres to a very different philosophy. Fibre Channel is most popular in storage networks end-to-end stability, reliability, and security are emphasized in favor over low cost and dynamic scalability. In Fibre Channel networks, the connecting ports must be fully authorized to communicate with their well-defined neighbors. Bandwidth for properly connected devices is tuned to avoid loss due to congestion. Also, routes for traffic are converged in advance, ensuring that only one route is used by any given traffic stream so that packets arrive in their expected sequence.

Ethernet and Fibre Channel networks are coming into contact with each other more frequently in modern networks. In some cases, legacy Fibre Channel devices are connected via Ethernet networks using Converged Enhanced Ethernet (CEE), a collection of recent Ethernet features designed to satisfy Fibre Channel delivery expectations. Although not the focus of this chapter, the CN4093 supports CEE and Fibre Channel over Ethernet (FCoE). For details, see "FCoE and CEE" on page 303.

Another approach is to use converged switches such as the CN4093 to support direct connection to both Ethernet and Fibre Channel networks. This allows a "best of both worlds" approach, using ubiquitous Ethernet networks for regular traffic, and full connection to Fibre Channel networks for lossless applications and the legacy architecture of established Storage Area Networks.

Supported Switch Roles

The CN4093 can be used in the following Fibre Channel applications:

- Node Port Virtualized gateway (NPV), linking multiple FCoE nodes to an upstream full fabric switch.
- FC-BB-5 compliant full fabric FC/FCoE switch.

These functions are independent of each other, and can coexist or be used in combination on the switch.

The CN4093 acts as a bridge between FCoE traffic and the Fibre Channel network. The switch performs Ethernet encapsulation for traffic heading to Ethernet ports, and performs decapsulation for traffic to Fibre Channel ports.

FCoE features are covered in "FCoE and CEE" on page 303. These features can be used independently or in conjunction with NPV gateway and full fabric switch features.

NPV Gateway

As a Node Port Virtualized (NPV) gateway, the CN4093 can act as a Fibre Channel collector, connecting numerous Fibre Channel end-point devices (known as nodes) for uplink to a Fibre Channel full fabric switch, performing stateless FC/FCoE encapsulation and decapsulation.

This helps resolve a typical problem in Fibre Channel networks where port density is low on Director Class SAN switches, or considered too valuable to relegate to individual nodes. As an NPV gateway, the CN4093 acts as a proxy to the upstream full fabric switch on behalf of the connected nodes.

The CN4093 supports standard Node Port Identifier Virtualization (NPIV) behavior.

The NPV gateway allows concurrent logins from multiple node ports (and multiple server connections) to be forwarded upstream to the full-fabric switch.

The upstream switch provides full fabric services such as zoning enforcement, and makes all switching decisions.

The gateway switch appears as a Fibre Channel end node to the full fabric switch, and acts as proxy for the full fabric switch to its connected node devices.

When multiple uplink ports are available between the NPV gateway and the upstream switch, nodes are not ensured to be assigned the same uplink whenever they request a session.

Only FCoE end nodes are allowed to be downstream connections. All Fibre Channel devices are connected to the full fabric FC switch.

Full-Fabric FC/FCoE Switch

As a full fabric FC/FCoE switch, the CN4093 authenticates connecting neighbors, provides Fibre Channel IDs, enforces port security among zones, and informs neighboring devices of network changes.

When acting as a full-fabric switch, the CN4093 can be connected to NPV gateways or directly to Fibre Channel nodes. In full-fabric mode, the CN4093 can be connected directly to another full fabric CN4093 or a *Lenovo RackSwitch G8264CS* through Fibre Channel ISL. For further details, see "E-Ports" on page 345.

Limitations

In Enterprise NOS 8.4, CN4093 does not support the following Fibre Channel port types:

- FL ports connecting storage fabric loop devices.

Implementing Fibre Channel

This section describes the basic elements of Fibre Channel configuration. For examples combining these elements, see "Fibre Channel Configuration" on page 348

Note: Use only the ISCLI or BBI to configure Fibre Channel. Lenovo N/OS CLI is not supported. After configuring Fibre Channel, save any subsequent configurations only in ISCLI or BBI. If Lenovo N/OS CLI is used to save any switch configuration, the Fibre Channel configuration will be lost.

Note: If you need to change the Fibre Channel configuration mode from BBI to ISCLI, first save the configuration using the `save` button on the BBI. Do not use two configuration modes at the same time.

Port Modes

The CN4093 has the following types of network port:

- **Ethernet Ports (Internal)**

  INTA1-INTA14 (ports 1-14), INTB1-INTB14 (15-28), INTC1-INTC14 (29-42)

  These standard 10Gb SFP+ Ethernet ports connect internally to servers in the system chassis.

- **Ethernet Ports (External**

  EXT1-EXT2 (ports 43-44)

  These standard 10Gb SFP+ Ethernet ports provide external connectors.

- **High-Capacity Ethernet Ports (External)**

  EXT3-EXT10 (ports 45-52)

  These 40Gb QSFP+ Ethernet ports can be configured as either two 40Gb Ethernet ports (EXT3 and EXT7), or as four 10Gb Ethernet ports (EXT3-EXT6, EXT7-EXT10).

- **Omni Ports (External)**

  EXT11-EXT22 (ports 53-64)

  These 10Gb SFP+ hybrid ports can be configured to operate either in Ethernet mode (the default) or in Fibre Channel mode for direct connection to Fibre Channel devices.

All Ethernet ports (including the Omni Ports by default) can carry any Ethernet data traffic, including Fibre Channel over Ethernet (FCoE) traffic.

When configured to operate in Fibre Channel mode, Omni Ports can be used as Fibre Channel downlinks (connected to Fibre Channel servers or storage devices) or as uplinks (connected to the data center SAN network).

The Omni Port mode (Ethernet or Fibre Channel) can be changed only for pairs of ports. The following port pairs share the same mode: EXT11-EXT12, EXT13-EXT14, EXT15-EXT16, EXT17-EXT18, EXT19-EXT20, EXT21-EXT22. Any combination of the mentioned port pairs is allowed for converting into FC ports.

Paired ports need not be connected to the same device and can even be used in different VLANs. The only required mutual attribute is the network type (Ethernet or Fibre Channel).

Fibre Channel configuration requires that at least one pair of Omni Ports be set to Fibre Channel mode.

The mode for Omni Port pairs or ranges can be configured using the following privileged configuration command:

```
CN 4093(config)# [no] system port <low port>-<high port> type fc
```

Fibre Channel VLANs

On the CN4093, each Fibre Channel network connected to the switch must be assigned its own VLAN. For each VLAN used with Fibre Channel, following properties must be defined:

- VLAN number
- Switch role (NPV mode or full fabric mode)
- Port membership
- Fibre Channel ports roles (as uplink ports or node connections)

The following commands are used to define a typical VLAN:

- Set or delete a VLAN

  ```
  CN 4093(config)# [no] vlan <VLAN number>
  ```

  FCoE networks typically use VLAN 1002. If using a different VLAN for FCoE, be sure that any connected servers and FCoE bridge will support your selection.

  This command initiates VLAN configuration mode. All VLAN-related Fibre Channel configuration is performed in this mode.

- Enable or disable the VLAN

  ```
  CN 4093(config-vlan)# [no] shutdown
  ```

- Exit VLAN configuration mode

  ```
  CN 4093(config-vlan)# exit
  ```

Port Membership

As with typical VLAN configuration, each VLAN used with a Fibre Channel network must include a description of its port members. VLANs used in Fibre Channel networks follow typical VLAN configuration rules (see "VLANs" on page 141), with the following additions:

- An FC port may belong to only one Fibre Channel VLAN.
- Trunking must be enabled for FCoE Ports and these ports must be part of the Fibre Channel VLAN.
- At least one Fibre Channel port must be included in each Fibre Channel VLAN.

From within VLAN configuration mode, the following command is used to add or remove port members:

```
CN 4093(config-vlan)# interface port <port number, alias or range>
CN 4093(config-if)# switchport trunk allowed vlan [add|remove] <VLAN ID>
```

For example:

```
CN 4093(config-vlan)# interface port INTA5,INTB5,INTC5
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan 1,1002
CN 4093(config-if)# exit

CN 4093(config)# interface fc EXT11, EXT12
CN 4093(config-if)# switchport trunk allowed vlan 1,1002
CN 4093(config-if)# exit
```

Notes:

- When using port aliases to specify a small range of internal ports, the aliases are converted to regular port numbers before applying the range. For example, specifying `inta5-intc5` (mixing A and C aliases) would result in a broad range than includes all ports between 5 (INTA5) and 33 (INTC5), and not just the lesser group of INTA5, INTB5, and INTC5.
- FC Omni ports must also be members of VLAN 1 (default VLAN), whereas FCoE ports can have any native VLAN. A Fibre Channel VLAN must not be configured as the native VLAN of the FCoE ports.

Switching Mode

The switch's role in the Fibre Channel network can be defined on a per-VLAN basis. The administrator can specify one of the following modes:

- NPV mode to uplink one or more nodes to a full fabric switch
- Full fabric mode

The CN4093 supports up to 12 Fibre Channel VLANs at any given time. Only one mode can be active on any specific VLAN at a given time, and only one VLAN can operate in full fabric mode.

From within VLAN configuration mode, the following commands are used to specify the Fibre Channel mode:

- To enable or disable NPV mode:

```
CN 4093(config-vlan)# [no] npv enable
```

- To enable or disable full fabric mode:

```
CN 4093(config-vlan)# [no] fcf enable
```

Note: Trunking must be enabled for FCoE Ports and these ports must be part of the Fibre Channel VLAN (NPV/FCF).

NPV Gateway

As a Node Port Virtualized (NPV) gateway, the CN4093 can act as a Fibre Channel collector, connecting numerous Fibre Channel end-point devices (known as nodes) for uplink to a Fibre Channel full fabric switch, performing stateless FC/FCoE encapsulation and decapsulation. For more details, see "NPV Gateway" on page 335.

NPV Port Traffic Mapping

Within each VLAN used with Fibre Channel, the physical ports may be used in the following roles:

- NPV External Interfaces

  All NPV gateways on the CN4093 must connect to a full fabric switch. The NPV external interface map specifies which Fibre Channel port or ports (Omni Ports set to Fibre Channel mode) are used for this purpose within each Fibre Channel VLAN. At least one Fibre Channel port is required, though two are typically used in order to provide redundancy. The following VLAN configuration command is used to define or remove the uplink:

  ```
  CN 4093(config-vlan)# [no] npv traffic-map external-interface <ports>
  ```

- Fibre Channel over Ethernet node

  Traffic from Ethernet ports which are properly configured to use CEE and FCoE (see "FCoE and CEE" on page 303) is permitted with no additional configuration.

- Ethernet

  Traffic on regular (non-FCoE) Ethernet ports will be blocked on Fibre Channel VLANs.

NPV Disruptive Load-Balancing

Every server connected to the NPV gateway logs into an upstream FC switch through a NP uplink. If multiple NP uplinks are available in a NPV VLAN, the logins are evenly distributed over the available uplinks.

The number of logins per uplink can go out of balance if a failed NP uplink is restored or a new uplink is brought online. The NPV gateway does not automatically move Enodes from the existing to new uplinks in such situations. To force the logins to be evenly distributed among all available uplinks in a NPV VLAN, the load-balancing CLI is available under the VLAN config.

Disruptive load-balancing can be configured through two available options:

- **manual** - triggers a disruptive load-balance among the logged-in nodes in the current NPV VLAN. The CLI for this option is:

  ```
  CN 4093(config-vlan)# npv disruptive-load-balance
  ```

  Note: This option addresses only the current imbalance. The CLI must be run again if any event causes uneven distribution in the future.

- **automated** - triggers a disruptive load-balance among the logged-in nodes in case of any future imbalances that are caused by Fibre Channel uplink flapping or introduction of new Fibre Channel uplink. If the Fibre Channel port is part of a NPV VLAN and is traffic-mapped, the load-balancing algorithm will assess the load on existing links and redistribute Enodes to the new uplink if necessary. The CLI for this option is:

```
CN 4093(config-vlan)# [no] npv auto-disruptive-load-balance enable
```

Note: This option triggers a disruptive load-balance 60 seconds after the FC port came online. If any uplinks are flapping in the configured VLAN, the disruptive load-balance will not be triggered.

Note: The automated option will only take care of the imbalances caused by Fibre Channel uplinks flapping, not Enodes flapping. In the later case, manual load-balance command should be used.

Note: To check which VLANs are have automated disruptive load-balancing enabled, use the following command:

```
CN 4093(config-vlan)# show npv auto-disruptive-load-balance
```

The load-balancing is disruptive in nature as few devices are forced to logout and initiate a re-login. The switch attempts to limit disruption by moving the fewest nodes necessary.

Limitations

Only Emulex CNAs participate in load-balancing. FCoE targets (Ex: V7000) are not load-balanced.

Other CNAs (such as Qlogic) store FCF information and try to login to the same FCF (uplink), so they are not balanced. Servers with AIX Operating Systems also can't be load-balanced for this reason.

Full Fabric Mode

As a full fabric FC/FCoE switch, the CN4093 authenticates connecting neighbors, provides Fibre Channel IDs, enforces port security among zones, and informs neighboring devices of network changes. For more details, see "Full-Fabric FC/FCoE Switch" on page 336.

Full Fabric Zoning

The CN4093 supports Fibre Channel zones and zonesets for VLANs operating in full fabric mode. In NPV gateway mode, zoning is controlled by the upstream full fabric switch and is not configurable in the NPV gateway VLAN.

Zoning allows logical grouping of ports and storage devices within a storage area network. Zoning defines access control between groups of servers and storage devices.

A SAN typically is divided into zones and zonesets, as described in the following sections.

Zones

A *zone* is a logical grouping of end nodes that are permitted to interact with each other. Zones can be grouped into a zonesets, which can be activated or deactivated as a single entity. A zone provides security by restricting access to only those devices that reside within the zone. Zoning also confines change notification floods within each zone.

Each zone contains one or more servers and one or more storage devices. Ports and devices in a zone are called zone members. A zone contains one or more zone members. A device can belong to one or more zones. End nodes that are members of a zone can communicate with each other, but they are isolated from nodes in other zones of which they are not a member.

Note: Only use the default zoneset with a limited number of FCoE connections or for testing purposes. If you need multiple FCoE connections, it is recommended that you configure well-defined zoning. If the default zone is configured to permit member services to communicate with each other, the fabric cannot have more than 20 logins.

If no zone is configured for the device, it resides in the default zone. You can configure the default zone to permit or deny its member devices to communicate with each other.

You can specify zone members based on any of the following criteria:

- **pWWN**: The port World Wide Number is a unique ID represting a particular end node. The pWWN is a 64-bit hexadecimal value (for example, 20:34:00:80:e5:23:f4:55)
- **FC ID**: The Fibre Channel identifier (FC ID) specifies the unique fabric domain ID of a device that connects to a node port on the switch. The FC ID is assigned by the full-fabric switch during the connection sequence and can change if the device logs out of the Fibre Channel fabric and returns. The FC ID is a 24-bit hexadecimal value (for example, 0xab00c1), but can also be specified in hexadecimal dotted notation (for example, ab.00.c1) generally representing *<domain or device>.<area>.<link>*

- **FC alias**: The Fibre Channel alias specifies the device that connects to a node port on the switch. The FC alias is a 1-64 character text value (for example, StorageOne).

  Note: When you create an FC alias using SNMP, a default pWWN of value 10:00:00:00:00:00:00:00 is automatically assigned to the FC alias. You must change this default pWWN value. Not doing so will result in a conflict of pWWN IDs the next time you try to create an FC alias.

Note: The CN4093 uses hard zoning, which is enforced in the switch hardware, based on the pWWNs of the Fibre Channel initiators and targets.

The CN4093 supports up the 64 zones per zoneset, each with up to 20 member devices. However, when an FC alias is used, only 10 devices can be members of a zone.

Zonesets

Zonesets provide a mechanism for conveniently grouping zones. Each zoneset can contain one or more zones. A zone can belong to one or more zonesets.

Only one zoneset can be activated at a given time. If you deactivate the active zoneset, no zonesets are active until you activate another zoneset. If you activate one zoneset while another zoneset is active, the currently active zoneset is deactivated.

When you activate a zoneset, the new zoneset access policies are applied.

Up to four zonesets can be configurated on the switch at any given time, though only one can be active.

Traffic flow between end devices is restricted by default setting.Its recommended that user should have well-defined zoning configuration activated for security concerns. To configure default-zone policy, use the following command:

```
CN 4093(config)# [no] zone default-zone permit
```

Defining Zoning

Define the following general properties for Fibre Channel zones and zonesets:

1. If desired, create (or remove) aliases for Fibre Channel devices:

   ```
   CN 4093(config)# [no] fcalias <device alias name> wwn <port World Wide Name>
   ```

   Repeat for each alias as necessary.

2. For each desired zone:

   a. Name (or remove) the zone.

      ```
      CN 4093(config)# [no] zone name <zone name>
      ```

   b. Add (or remove) one or more members to the zone using either pWWNs, FC IDs, or FC Aliases

      ```
      CN 4093(config-zone)# [no] member pwwn <port World Wide Name>
      CN 4093(config-zone)# [no] member fcid <Fibre Channel ID>
      CN 4093(config-zone)# [no] member fcalias <device alias name>
      ```

Repeat as necessary for each member device.

c. Exit from zone configuration:

```
CN 4093(config-zone)# exit
```

3. For each desired zoneset:

a. Name (or remove) the zoneset.

```
CN 4093(config)# [no] zoneset name <zoneset name>
```

b. Add (or remove) one or more member zones to the zoneset:

```
CN 4093(config-zoneset)# [no] member <zone name>
```

Repeat as necessary for each member zone.

c. Exit from zoneset configuration:

```
CN 4093(config-zone)# exit
```

Activating Zoneset

Fibre Channel is intended to operate with minimal disruption. To prevent the various synchronization events that would result if each stage of a live zoning configuration was applied, the cumulative configuration changes for zones and zoneset are held in reserve until explicitly activated by the administrator.

When activated, the new zoneset will be synchronized throughout the Fibre Channel fabric for each modified zone. Fibre Channel traffic will be temporarily disrupted in modified zones as changes to the fabric are recognized by the connected devices. Until activation, the previously established zoneset will remain in effect.

The basic zoneset commands are as follows:

- Activate or deactivate a zoneset:

```
CN 4093(config)# [no] zoneset activate name <zoneset name>
```

- View the settings for the active zoneset:

```
CN 4093# show zoneset active
```

- View the settings for the pending configuration changes:

```
CN 4093# show zoneset
```

E-Ports

E-ports (expansion ports) connect two full fabric switches to form an inter-switch link (ISL). Up to four Fibre Channel ISLs can be established between two full fabric switches.

Only Fibre Channel port types can be configured as E-ports. These ports must be members of a Fibre Channel VLAN. Use the following commands to configure E-ports:

```
CN 4093(config)# system port <port range> type fc
CN 4093(config)# interface fc <port range>
CN 4093(config-if)# type e
```

To verify the port configuration and operational state, use the following command:

```
CN 4093(config)# show interface fc information
```

In a stack of switches, ports on the Master or Backup switches can be configured as E-ports.

If different zones were configured on the switches being connected using E-ports, the zones merge to establish a consistent zoning policy across the fabric. The zones on the two switches must belong to any one zoneset.

Note: Zones are added in the merged zoneset only if the zoneset on the individual switch was active.

The following table lists the zone merge rules:

Table 31. *Zone merge rules*

Adjacent Zoning Configuration	Local Zoning Configuration	Result in Local Switch
Zone Set State is Deactivated.	Zone Set State is Activated or Deactivated.	No change
Zone Set State is Activated.	Zone Set State is Deactivated.	Zone Set gets the Adjacent Zone Set State (i.e., the Zone Set is activated). Zone Set gets the adjacent Zone Set.
Adjacent Zone Set is equal to the Local Zone Set		No change
Adjacent and Local Zone Sets contain a Zone with the same name but different members.		ISL Isolated.
Adjacent Zone Set contains Zones that are not included in the Local Zone Set, and/or Local Zone Set contains Zones that are not included in the Adjacent Zone Set.		Zone Set State becomes Activated. Zone Set is the merge of the local Zones plus the Adjacent Zones.

E-ports cannot be used to form stack LAG links.

- Enterprise NOS supports ISL distance up to 3 kms.
- E-ports can be configured only on Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch and Lenovo RackSwitch G8264CS. E-ports cannot interoperate with the switches from other vendors.
- A maximum of eight switches are supported in a Fibre Channel fabric.
- A maximum of four E-ports can be configured on a switch.

Optimized FCoE Traffic Flow

To optimize FCoE traffic flow between FCoE enodes, optimized-forwarding feature installs appropriate ACL entries for logged-in nodes. Usually, any FC/FCoE traffic in full fabric mode should go through FC module for Zone check. We can achieve low latency if the Zone check is done on Ethernet switch module for FCoE-FCoE traffic. Optimized feature is enabled by default in Full fabric mode, and is not applicable to NPV mode.

Note: FCoE-FC and FC-FC traffic is not optimized.

If needed, the administrator can disable optimized-forwarding feature. Prior to that, disable FIP snooping. Use the following commands:

```
CN 4093(config)# no fcoe fips enable
CN 4093(config)# no fcoe optimized-forwarding enable
```

To re-enable optimized-forwarding feature, use the following command sequence:

```
CN 4093(config)# no fcoe fips enable
CN 4093(config)# fcoe optimized-forwarding enable
CN 4093(config)# fcoe fips enable
```

To view optimized traffic flow information, use the following commands:

```
CN 4093(config)# show fcoe optimized-forwarding status
(Show current state of feature)

CN 4093(config)# show fcoe optimized-acls vlan <vlan ID>
(Show list of optimized ACLs)
```

Storage Management Initiative Specification (SMI-S)

Enterprise NOS provides a programming interface using the SMI-S to ease interoperability in a multivendor SAN environment. In this release, only limited support is provided. The CN4093 switch must be operating in full fabric mode.

An embedded SMI-S agent runs on the CN4093 and includes standard profiles as specified in the SMI-S. These profiles include:

- Fabric Profile
- Indication Profile
- Server Profile
- Switch Profile

Configuration capabilities of switch, fabric, or ports is not supported. Zoning control can be implemented and includes the following functions:

- Session control (start, commit, and rollback)
- Zoning updates
- Create and destroy zone set, zone, and zone alias
- Add/Remove zone to zone set, zone alias, or port WWN to zone and port WWN to zone alias
- Activate and deactivate zoneset

The IBM Director (includes Tivoli Storage Productivity Center (TPC)) is used to configure and administer the fibre channel fabric. Connection with the SMI-S agent can be established via IPv4 or IPv6 management interface using HTTP/HTTPS. Use the following link:

```
http://<Management IP address>:5988
(OR)
https://<Management IP address>:5989
```

The namespace for the SMI-S agent is root/interop.

You will need to authenticate using the login and password configured for the switch.

Restrictions

The current implementation of SMI-S does not support the following:

- NPV mode
- Zone configuration on the switch that uses FC IDs instead of port WWNs.

  Note: The CLI commands may be available, but the zone configuration will not be applied.

Fibre Channel Configuration

Configuration Guidelines

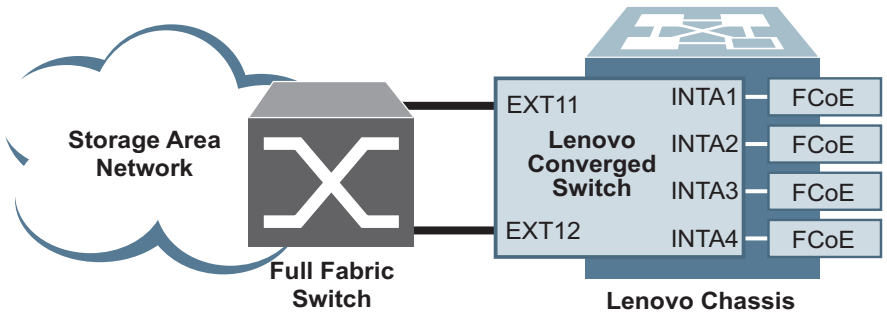
In Enterprise NOS 8.4, consider the following when configuring Fibre Channel on the CN4093:

- Up to 12 Fibre Channel networks (on separate VLANs) are supported.
- Only one Fibre Channel VLAN can operate in full fabric mode at any given time. All others can operate as NPV gateways.
- Only Omni Ports (EXT11-EXT22) can be placed in Fibre Channel mode. All other ports operate in Ethernet mode, which may participate in Fibre Channel networks as FCoE nodes.
- At least one Omni Port is required to operate in Fibre Channel mode in each Fibre Channel VLAN.
- Zones and zonesets apply only to a VLAN in full fabric mode. Up to 4 zonesets may be configured, but only 1 can be active at any given time. The CN4093 supports up the 64 zones per zoneset, each with up to 20 member devices. However, when an FC alias is used, only 10 devices can be members of a zone.
- In a stack setup, the full fabric mode or NPV gateway mode is bound to a single switch, as determined by the external FC ports in the VLAN. All FC ports belonging to a Fibre Channel VLAN must be part of the same switch. Both full fabric mode and NPV mode can be configured on the Master as well as the Backup switch. For details on configuring a stack, see Chapter 13, "Stacking".

Example 1: NPV Gateway

In this example, the CN4093 operates as an NPV gateway:

Figure 35. Using the CN4093 as an NPV Gateway



The switch connects to FCoE node ports to an external Fibre Channel full fabric switch. Because multiple nodes will share the CN4093 uplinks, the network must be configured as an NPV gateway.

Note: Up to 12 Fibre Channel VLANs can be configured on the switch at any given time, any or all of which can be configured as NPV gateways.

1. Specify which Omni Ports are directly connected to Fibre Channel devices:

```
CN 4093(config)# system port ext11-ext12 type fc
```

Note: On the CN4093, FC devices can be connected only to Omni Ports. Omni Ports connected to FCoE devices are considered part of the Ethernet network and should be left to operate in Ethernet mode.

2. Enable tagging/trunk mode for internal ports participating in FCoE:

```
CN 4093(config)# interface port inta1-inta4
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit
```

3. Specify all member ports for the VLAN:

```
CN 4093(config)# interface port inta1-inta4
CN 4093(config-if)# switchport trunk allowed vlan add 1002
CN 4093(config-if)# exit
CN 4093(config)# interface fc ext11-ext12
CN 4093(config-if)# switchport trunk allowed vlan add 1002
CN 4093(config-if)# exit
```

4. Specify a VLAN for this Fibre Channel network:

```
CN 4093(config)# vlan 1002
```

5. Enable NPV mode on the VLAN:

```
CN 4093(config-vlan)# npv enable
```

6. Specify which external ports are connected to the upstream Fibre Channel full fabric switch:

```
CN 4093(config-vlan)# npv traffic-map external-interface ext11-ext12
CN 4093(config-vlan)# exit
```

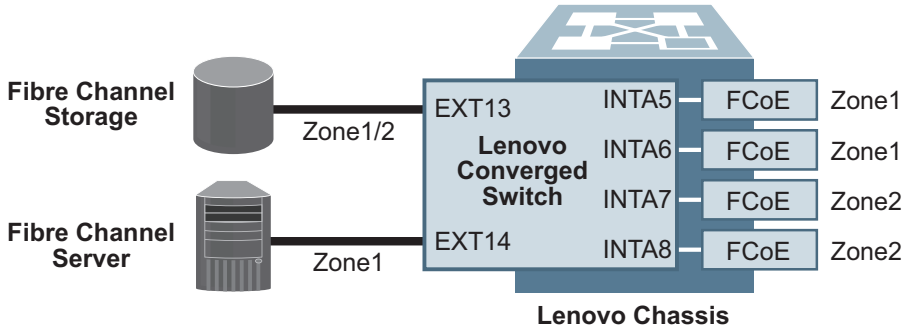
Note: Although this example depicts two Fibre Channel ports connected to the upstream device, this is done for the sake of network redundancy. Only one Fibre Channel port is actually required.

7. Remove unused ports—ports that are not part of the uplink to the Fibre Channel fabric—from the NPV VLAN.

Example 2: Full Fabric FC/FCoE Switch

Consider the following Fibre Channel network:

Figure 36. Using the CN4093 as a Full Fabric FC/FCoE Switch



In this example network, the CN4093 acts as the full fabric switch for the Fibre Channel network in two zones.

Note: Although up to 12 Fibre Channel VLANs can be configured on the switch at any given time, only one can operate in full fabric mode. The rest may be configured as NPV gateways. For instance, the full fabric configuration in this example can be used simultaneously with up to 11 NPV gateways configured as shown in the NPV example on page 350.

1. Specify which Omni Ports will be used for Fibre Channel devices:

```
CN 4093(config)# system port ext13-ext14 type fc
```

Note: On the CN4093, Fibre Channel devices can be connected only to Omni Ports. Omni Ports connected to FCoE devices are considered part of the Ethernet network and should be left to operate in Ethernet mode.

2. Enable tagging/trunk mode for internal ports participating in FCoE:

```
CN 4093(config)# interface port inta5-inta8
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# exit
```

3. Specify all member ports for the VLAN:

```
CN 4093(config)# interface port inta5-inta8
CN 4093(config-if)# switchport trunk allowed vlan add 200
CN 4093(config-if)# exit
CN 4093(config)# interface fc ext13-ext14
CN 4093(config-if)# switchport trunk allowed vlan add 200
CN 4093(config-if)# exit
```

Note: At least one Fibre Channel port must be included.

4. Specify a VLAN for the this Fibre Channel network:

```
CN 4093(config)# vlan 200
```

5. Enable full fabric mode on the VLAN:

```
CN 4093(config-vlan)# fcf enable
CN 4093(config-vlan)# exit
```

6. Define Fibre Channel zones:

```
CN 4093(config)# zone name Zone1
CN 4093(config-zone)# member pwwn 20:34:00:80:e5:23:b1:55
CN 4093(config-zone)# member pwwn 20:34:00:80:e5:27:f4:56
CN 4093(config-zone)# member pwwn 20:34:00:80:e5:28:31:13
CN 4093(config-zone)# member pwwn 20:34:00:80:e5:28:31:14
CN 4093(config-zone)# exit

CN 4093(config)# zone name Zone2
CN 4093(config-zone)# member pwwn 20:34:00:80:e5:28:43:57
CN 4093(config-zone)# member pwwn 20:34:00:80:e5:18:b3:58
CN 4093(config-zone)# member pwwn 20:34:00:80:e5:28:31:13
CN 4093(config-zone)# exit

CN 4093(config)# zoneset name City1
CN 4093(config-zoneset)# member Zone1
CN 4093(config-zoneset)# member Zone2
CN 4093(config-zoneset)# exit

CN 4093(config)# zoneset activate name City1
```

Fibre Channel Standard Protocols Supported

Following table lists the standard FC protocols supported on the CN4093 10 Gb Converged Scalable Switch.

Table 32. *FC Protocols Supported*

Protocol
Fibre Channel FCoE · T11 FCoE Initialization Protocol (FIP) (FC-BB-5) Fibre Channel forwarding (FCF)
Port Types Fibre Channel: F, NP, VF E (N/OS 7.8 onwards)
Sixteen Buffer credits supported:
Fabric Device Management Interface (FDMI)
NPIV
NPV Gateway
Fabric Shortest Path First (FSPF)
Port security
Fibre Channel ping, debugging
Fibre Channel Standards · FC-PH, Revision 4.3 (ANSI/INCITS 230-1994) FC-PH, Amendment 1 (ANSI/INCITS 230-1994/AM1 1996) FC-PH, Amendment 2 (ANSI/INCITS 230-1994/AM2-1999) FC-PH-2, Revision 7.4 (ANSI/INCITS 297-1997) FC-PH-3, Revision 9.4 (ANSI/INCITS 303-1998) FC-PI, Revision 13 (ANSI/INCITS 352-2002) FC-PI-2, Revision 10 (ANSI/INCITS 404-2006) FC-PI-4, Revision 7.0 FC-FS, Revision 1.9 (ANSI/INCITS 373-2003) FC-FS-2, Revision 0.91 FC_FS_3 Revision 1.11 FC-LS, Revision 1.2 FC-SW-2, Revision 5.3 (ANSI/INCITS 355-2001) FC-SW-3, Revision 6.6 (ANSI/INCITS 384-2004) FC-SW-5, Revision 8.5 (ANSI/INCITS 461-2010) FC-GS-3, Revision 7.01 (ANSI/INCITS 348-2001) FC-GS-4, Revision 7.91 (ANSI/INCITS 387-2004) FC-GS-6 Revision 9.4 (ANSI/INCITS 463-2010) FC-BB-5, Revision 2.0 for FCoE FCP, Revision 12 (ANSI/INCITS 269-1996) FCP-2, Revision 8 (ANSI/INCITS 350-2003) FCP-3, Revision 4 (ANSI/INCITS 416-2006) FC-MI, Revision 1.92 (INCITS TR-30-2002, except for FL-ports and Class 2) FC-MI-2, Revision 2.6 (INCITS TR-39-2005) FC-SP, Revision 1.6 FC-DA, Revision 3.1 (INCITS TR-36-2004)

Chapter 19. Edge Virtual Bridging

The 802.1Qbg/Edge Virtual Bridging (EVB) is an emerging IEEE standard for allowing networks to become virtual machine (VM)-aware. EVB bridges the gap between physical and virtual network resources. The IEEE 802.1Qbg simplifies network management by providing a standards-based protocol that defines how virtual Ethernet bridges exchange configuration information. In EVB environments, virtual NIC (vNIC) configuration information is available to EVB devices. This information is generally not available to an 802.1Q bridge.

Enterprise NOS EVB features are compliant with the IEEE 802.1Qbg Authors Group Draft 0.2. For a list of documents on this feature, see: http://www.ieee802.org/1/pages/802.1bg.html.

Enterprise NOS implementation of EVB supports the following protocols:

- Virtual Ethernet Bridging (VEB) and Virtual Ethernet Port Aggregator (VEPA): VEB and VEPA are mechanisms for switching between VMs on the same hypervisor. VEB enables switching with the server, either in the software (vSwitch), or in the hardware (using single root I/O virtualization capable NICs). VEPA requires the edge switch to support "Reflective Relay"— an operation where the switch forwards a frame back to the port on which it arrived if the destination MAC address is on the same port.
- Edge Control Protocol (ECP): ECP is a transport protocol that operates between two peers over an IEEE 802 LAN. ECP provides reliable, in-order delivery of ULP (Upper Layer Protocol) PDUs (Protocol Data Units).
- Virtual Station Interface (VSI) Discovery and Configuration Protocol (VDP): VDP allows hypervisors to advertise VSIs to the physical network. This protocol also allows centralized configuration of network policies that will persist with the VM, independent of its location.
- EVB Type-Length-Value (TLV): EVB TLV is a component of Link Layer Discovery protocol (LLDP)-based TLV used to discover and configure VEPA, ECP, and VDP.

EVB Operations Overview

The ENOS includes a pre-standards VSI Type Database (VSIDB) implemented through the System Networking Switch Center (SNSC), the Lenovo Flex System Manager (FSM), or the Lenovo System Networking Distributed Switch 5000V. The VSIDB is the central repository for defining sets of network policies that apply to VM network ports. You can configure only one VSIDB.

Note: This document does not include the VSIDB configuration details. Please see the SNSC, FSM, or Lenovo System Networking Distributed Switch 5000V guide for details on how to configure VSIDB.

The VSIDB operates in the following sequence:

1. Define VSI types in the VSIDB. The VSIDB exports the database when the CN4093 sends a request.
2. Create a VM. Specify VSI type for each VM interface. See the SNSC, FSM, or Lenovo System Networking Distributed Switch 5000V guide for details on how to specify the VSI type.

   The hypervisor sends a VSI ASSOCIATE, which contains the VSI type ID, to the switch port after the VM is started. The switch updates its configuration based on the requested VSI type. The switch configures the per-VM bandwidth using the VMpolicy.

   The Enterprise NOS supports the following policies for VMs:

   - ACLs
   - Bandwidth metering

VSIDB Synchronization

The switch periodically checks for VSIDB changes based on the configured interval. You can configure this interval using the following command:

```
CN 4093(config)# virt evb vsidb <number>
CN 4093(conf-vsidb)# update-interval <time in seconds>
```

To disable periodic updates, use the following command:

```
CN 4093(conf-vsidb)# no update-interval
```

If the switch finds that the VSIDB has changed, it updates the local VSIDB cache. When the cache is successfully updated, it sends a syslog message.

After updating the local VSIDB cache, the switch disassociates any VM whose type ID or VLAN no longer exists in the updated cache.

The switch updates the local VSIDB cache when any of the following takes place:

- When, at the configured refresh interval, the switch finds that the VSIDB configuration has changed since the last poll.
- When a VM sends an ASSOCIATE message, but the VSI type does not exist in the local VSIDB cache.
- When a VM sends an ASSOCIATE message, and the VSI type exists but the VSI type's VLAN ID does not exist in the local VSIDB cache.
- When you update the VSIDB using the following command:
  `CN 4093# virt evb update vsidb` *<number>*
- When the management port link status changes from down to up.

VLAN Behavior

When a VM gets associated, the corresponding VLAN is dynamically created on the switch port if the VLAN does not already exist.

VLANs that are dynamically created will be automatically removed from the switch port when there are no VMs using that VLAN on the port.

Dynamic VLAN information will not be displayed in the running configuration. However, the VLAN, port, and STP commands display the dynamic VLAN information with a "*".

If you configure any Layer 2/Layer 3 features on dynamically created VLANs, the VLAN information is displayed in the running configuration.

Deleting a VLAN

If you delete a VLAN that has a VM associated with it, you will see a warning message similar to the following:

```
Warning: Vlan 10 is used by VM and can't be removed.
```

The VMs will not get disassociated.If a VM is associated with a port, and you remove this port from a VLAN, you will see a warning message similar to the following:

```
Warning: Port INTB1 in Vlan 10 is used by VM and can't be removed.
```

The VMs will not get disassociated.

Manual Reflective Relay

Reflective Relay (RR) is an operation where the switch forwards a frame back to the port on which it arrived if the destination MAC address is on the same port. When an EVB profile is configured on a port, RR is automatically enabled on the port after capability exchange with the peer, using the IEEE802.1QBG protocol. This is the usual mode of operation.

When the switch interoperates with devices that do not support IEEE 802.1QBG protocols, RR can be manually configured using the following command:

```
CN 4093(config-if)# reflective-relay force
```

Manual RR and EVB profile cannot be configured on a port at the same time.

Note: If a port is a member of an isolated VLAN, the manual reflective relay will not work. See "Private VLANs" on page 157 for more information on isolated VLANs.

EVB Configuration

This section includes the steps to configure EVB based on the following values:

- Profile number: 1
- Port number: 1
- Retry interval: 8000 milliseconds
- VSI Database:
  - Manager IP: 172.31.37.187
  - Port: 80

Note: VSI Database can be accessed via HTTP or HTTPS. The manager IP can be configured with an IPv4 or IPv6 address.

1. Create an EVB profile.

```
CN 4093(config)# virt evb profile 1                    (Enter number from 1-16)
```

2. Enable Reflective Relay.

```
CN 4093(conf-evbprof)# reflective-relay
```

3. Enable VSI discovery.

```
CN 4093(conf-evbprof)# vsi-discovery
CN 4093(conf-evbprof)# exit
```

4. Add EVB profile to port.

```
CN 4093(config)# interface port 1
CN 4093(config-if)# evb profile 1                      (Enter EVB profile ID)
CN 4093(config-if)# exit
```

5. Configure ECP retransmission interval.

```
CN 4093(config)# ecp retransmit-interval 8000
(Enter retransmission interval in milliseconds (100-9000))
```

6. Set VSI database information.

```
CN 4093(config)# virt evb vsidb 1
CN 4093(conf-vsidb)# protocol {http|https}
(Select VSI database protocol; default is HTTP)
CN 4093(conf-vsidb)# host 172.31.37.187 [data-port|extm-port|mgt-port]
(Set VSI database Manager IP)
CN 4093(conf-vsidb)# port 80                    (Set VSI database Manager port)
CN 4093(conf-vsidb)# filepath "vsidb"           (Set VSI database document path)
CN 4093(conf-vsidb)# filename "all.xml"         (Set VSI database file name)
CN 4093(conf-vsidb)# update-interval 30         (Set update interval in seconds)
CN 4093(conf-vsidb)# exit
```

Note: When you connect to a SNSC VSIDB, the port/docpath configuration is as follows:

HTTP:

- Port: 40080
- Docpath: snsc/rest/vsitypes
- HTTPS:
- Port: 40443
- Docpath: snsc/rest/vsitypes
- When you connect to a 5000v VSIDB, the port/docpath configuration is as follows:
- Port: 80
- Docpath: vsitypes

Configuring EVB in Stacking Mode

A *stack* is a group of up to eight CN4093 10 Gb Converged Scalable Switch switches with Enterprise NOS that work together as a unified system. The switches in a stack are interconnected by a stack LAG in a local ring topology.

An operational stack must contain one Master and one or more Members, as follows:

- **Master**

  One switch controls the operation of the stack and is called the Master. The Master provides a single point to manage the stack. A stack must have one and only one Master. The firmware image, configuration information, and run-time data are maintained by the Master and pushed to each switch in the stack as necessary.

- **Member**

  Member switches provide additional port capacity to the stack. Members receive configuration changes, run-time information, and software updates from the Master.

- **Backup**

  One member switch can be designated as a Backup to the Master. The Backup takes over control of the stack if the Master fails. Configuration information and run-time data are synchronized with the Master.

For details on implementing the stacking feature, see "Stacking" on page 235.

EVB can be configured on any port in the stack. Use the Master to configure EVB on a port in the stack. The port numbers in a stack use the following format:

<switch number>:<port number>

Configure VSIDB on a data or management port that resides on the Master.

The Master process the EVB-related information for all the switch ports in a stack. The Master performs the VSIDB synchronization (See "VSIDB Synchronization" on page 354). The Master synchronizes all EVB changes with the Backup.

If the Master fails, the Backup takes over control of the stack. The VSI associations on the Master ports are lost. All other VSI associations remain unchanged.

Limitations

- If both ACL and egress bandwidth metering are enabled, traffic will first be matched with the ACL and will not be limited by bandwidth metering.
- ACLs based on a source MAC or VLAN must match the source MAC and VLAN of the VM. If not, the policy will be ignored and you will see the following warning message:

```
"vm: VSI Type ID 100 Associated mac 00:50:56:b6:c0:ff on port 6,
ignore 1 mismatched ACL"
```

Unsupported features

The following features are not supported on ports configured with EVB:

- LAG/VLAG
- vNIC
- VMready

Chapter 20. Static Multicast ARP

The Microsoft Windows operating system includes the Network Load Balancing (NLB) technology that helps to balance incoming IP traffic among multi-node clusters. In multicast mode, NLB uses a shared multicast MAC address with a unicast IP address. Since the address resolution protocol (ARP) can map an IP address to only one MAC address, port, and VLAN, the packet reaches only one of the servers (the one attached to the port on which the ARP was learnt).

To avoid the ARP resolution, you must create a static ARP entry with multicast MAC address. You must also specify the list of ports through which the multicast packet must be sent out from the gateway or Layer 2/Layer 3 node.

With these configurations, a packet with a unicast IPv4 destination address and multicast MAC address can be sent out as per the multicast MAC address configuration. NLB maps the unicast IP address and multicast MAC address as follows:

Cluster multicast MAC address: 03-BF-W-X-Y-Z; where W.X.Y.Z is the cluster unicast IP address.

You must configure the static multicast ARP entry only at the Layer 2/Layer 3 or Router node, and not at the Layer 2-only node.

Enterprise NOS supports a maximum of 20 static multicast ARP entries.

Note: If you use the ACL profile or IPMC-OPT profile, an ACL entry is consumed for each Static Multicast ARP entry that you configure. Hence, you can configure a maximum of 640 ACL and multicast MAC entries together.The ACL entries have a higher priority. In the default profile, the number of static multicast ARP entries that you configure does not affect the total number of ACL entries.

Configuring Static Multicast ARP

To configure multicast MAC ARP, you must perform the following steps:

- Configure the static multicast forwarding database (FDB) entry: Since there is no port list specified for static multicast ARP, and the associated MAC address is multicast, you must specify a static multicast FDB entry for the cluster MAC address to limit the multicast domain. If there is no static multicast FDB entry defined for the cluster MAC address, traffic will not be forwarded. Use the following command:

```
CN 4093(config)# mac-address-table multicast <cluster MAC address> <port(s)>
```

- Configure the static multicast ARP entry: Multicast ARP static entries should be configured without specifying the list of ports to be used. Use the following command:

```
CN 4093(config)# ip arp <destination unicast IP address> <destination multicast MAC address> vlan <cluster VLAN number>
```

Configuration Example

Consider the following example:

- Cluster unicast IP address: 10.10.10.42
- Cluster multicast MAC address: 03:bf:0a:0a:0a:2a
- Cluster VLAN: 42
- List of individual or port LAGs to which traffic should be forwarded: 54 and 56

Following are the steps to configure the static multicast ARP based on the given example:

1. Configure the static multicast FDB entry.

```
CN 4093(config)# mac-address-table multicast 03:bf:0a:0a:0a:2a 42 54,56
```

2. Configure the static multicast ARP entry:

```
CN 4093(config)# ip arp 10.10.10.42 03:bf:0a:0a:0a:2a vlan 42
```

You can verify the configuration using the following commands:

- Verify static multicast FDB entry:

```
CN 4093(config)# show mac-address-table multicast address
03:bf:0a:0a:0a:2a

 Multicast Address  VLAN  Port(s)
 -----------------  ----  ---------
 03:bf:0a:0a:0a:2a    42  54 56
```

- Verify static multicast ARP entry:

```
CN 4093(config)# show ip arp

Mgmt ARP entries:

Total number of Mgmt ARP entries : 3
    IP address    Flags    MAC address     VLAN   Age Port
  --------------- ----- ----------------- ------  --- ----
  10.241.38.1           00:11:25:c3:70:0a   4095    1 MGT1
  10.241.38.101         00:11:25:c3:70:0a   4095    2 MGT1
  10.241.38.102    P    74:99:75:08:9b:ef   4095      MGT1

Data ARP entries:

Current ARP configuration:
 rearp 5

Current static ARP:
    IP address       MAC address    Port   VLAN
  --------------- ----------------- ----   ----
  10.10.10.42     03:bf:0a:0a:0a:2a          42

Total number of arp entries : 2
    IP address    Flags    MAC address    VLAN  Age Port
  --------------- ----- ----------------- ----  --- ----
  10.10.10.1       P    fc:cf:62:9d:74:00   42
  10.10.10.42      P    03:bf:0a:0a:0a:2a   42      0
```

Limitations

- You must configure the ARP only in the Layer 2/Layer 3 node or the router node but not in the Layer 2-only node. Enterprise NOS cannot validate if the node is Layer 2-only.
- The packet is always forwarded to all the ports as specified in the Multicast MAC address configuration. If VLAN membership changes for the ports, you must update this static multicast MAC entry. If not, the ports, whose membership has changed, will report discards.
- ACLs take precedence over static multicast ARP. If an ACL is configured to match and permit ingress of unicast traffic, the traffic will be forwarded based on the ACL rule, and the static multicast ARP will be ignored.

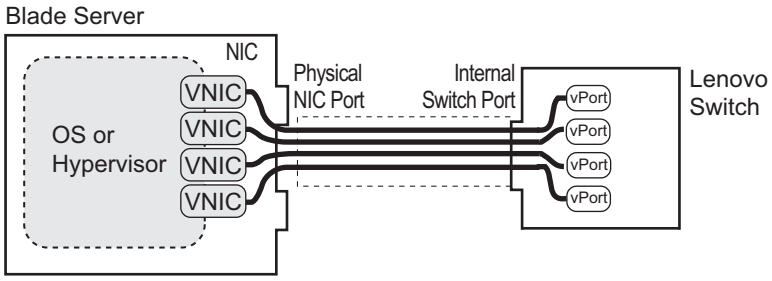
Chapter 21. Unified Fabric Port

Unified Fabric Port (UFP) is a cost-effective way to allocate, share and dynamically control network bandwidth between a server and a switch. UFP lets you create multiple virtual connections. The UFP protocol is a link-level protocol that runs a separate instance for each physical communication link established between a server NIC and a switch port. Virtualizing the ports allows you to separate or aggregate port traffic by applying the network policies defined on the switch. Virtualization lessens bottlenecks and provides higher bandwidth while consolidating equipment use.

UFP provides a switch fabric component to control NICs. The server operating system (OS) or hypervisor recognizes each subdivided link (channel) as an independent physical NIC. Each channel has a unique identity and profile that defines its properties and functionality. The server communicates with the switch over the channel as defined in the channel profile. The channels share the high-speed physical link bandwidth.

For each channel, the vNIC on the server side communicates with virtual port on the switch side. Any 10 Gbps internal server port can be configured as a UFP port.

Figure 37. UFP vPorts



The UFP protocol has the following operation categories:

- Channel Initialization: The server NIC and the switch port negotiate the number of channels and establish channel identifiers. Each UFP channel has a data component and a control component. The two components have the same UFP channel ID.
- Channel Control: For an established channel, the switch can modify configurable channel properties by sending a control message on the UFP channel. While the channel ID is the same for the control and data components, the destination MAC address of the control message frame is a well-known address `01-80-C2-00-00-03`.
- Discovery Capability: UFP can discover other ports that are UFP enabled. Once you enable UFP, you can check the information statistics for established channels.

UFP Limitations

The following limitations apply when configuring UFP:

- FCoE must be configured only on vPort 2 of the physical NIC for Emulex CNA and on vPort 1 of the physical NIC for Qlogic CNA. For Emulex NIC CN4052, FCoE can be configured on vPort 2 or vPort 3.
- UFP port in FCoE mode cannot operate with FIP auto-VLAN feature.
- VLANs that have member vPorts configured in trunk, access or auto modes cannot have member vPorts configured in tunnel mode or FCoE.
- vPorts on a physical port must be members of separate VLANs.
- VLANs 4002-4009 are reserved for outer tagging.
- A UFP-enabled port with no vPorts configured cannot belong to the same VLAN as a UFP-enabled port that has vPorts configured in trunk, access or auto modes.
- UFP bandwidth is guaranteed lossless only for unicast traffic.
- VMready is supported only on a vPort which is configure in auto-VLAN mode. When a vPort is in auto-VLAN mode, it can support up to 32 VMGroups.
- EVB is supported only on a vPort which is configured in auto-VLAN mode.
- VMready and EVB cannot be configured on the same physical port.
- UFP vPorts support up to 1024 VLANs in trunk and auto mode on the switch in standalone mode. Stacking switches have a limitation of 256 VLANs in both auto and trunk mode.
- When CEE is turned on, FCoE vPort must be used for lossless priority traffic. For loss-tolerant priority traffic, a non-FCoE UFP vPort must be used. The lossless property of FCoE vPort is not guaranteed, if lossless and loss-tolerant traffic are combined.
- When the vPort is enabled and the channel link state is up, the system does not support changing vPort VLAN type from private/non-private to non-private/private.
- A maximum of eight vPorts can be configured for each physical switch port. QoS Enhanced Transmission Selection (ETS) mode must be enabled when configuring more than four vPorts of a physical switch port. For more details about ETS mode, check page 370.
- UFP and vNIC cannot be configured enabled at the same time on a switch.
- VMReady Local Group configuration is not supported by UFP.
- If QoS ETS mode is used, a FCoE vPort must be configured with priority 3.
- UFP vPorts cannot be aggregated to form a LAG/vLAG client.

Virtual Ports Modes

A single physical switch port is configured with virtual ports (vPorts). Each UFP channel connects the server vNIC with a switch vPort. Properties that are defined for a vPort, such as native VLAN and bandwidth, are applied to the traffic that belongs to the vPort.

Note: A maximum of eight vPorts can be configured for each physical switch port. QoS Enhanced Transmission Selection (ETS) mode must be enabled when configuring more than four vPorts of a physical switch port. For mode details about ETS mode, check page 370.

vPort-S-Tag Mapping

A vPort can also be identified with an S-tag (service tag or outer tag). When a vPort is initialized, the switch communicates the UFP channel ID of the vPort to the server vNIC. When the server NIC or switch transmit frames, they add this S-tag to indicate the vPort or vNIC to which the packet is being transmitted. No VLAN mapping is required. Such packets can be single tagged or double tagged (with S-tag).

vPort-VLAN Mapping

In local domain data path type, the switch and server identify the vPort and vNIC by the port and VLAN tag in the incoming and outgoing packets. Because no two vPorts carry traffic for the same VLAN, the port-and-VLAN combination must be uniquely mapped to a vPort.

UFP vPort Mode

The UFP mode is configured based on the type of switching domain (single VLAN or multiple VLANs) where the vPort is connected.

- Use local domain data path types for trunk or access mode.
- Use pass-through domain data path types for tunnel mode. In tunnel mode, a vPort can belong to only one VLAN.

Use the following command to configure UFP vPort mode:

```
CN 4093(config)# ufp port <num> vport <num>
CN 4093(config_ufp_vport)# network mode {access|trunk|auto|tunnel|fcoe}
```

Notes:

- Default mode is `tunnel`.
- If LACP is down on a UFP port, the FCoE connections go down and are then reestablished by UFP.

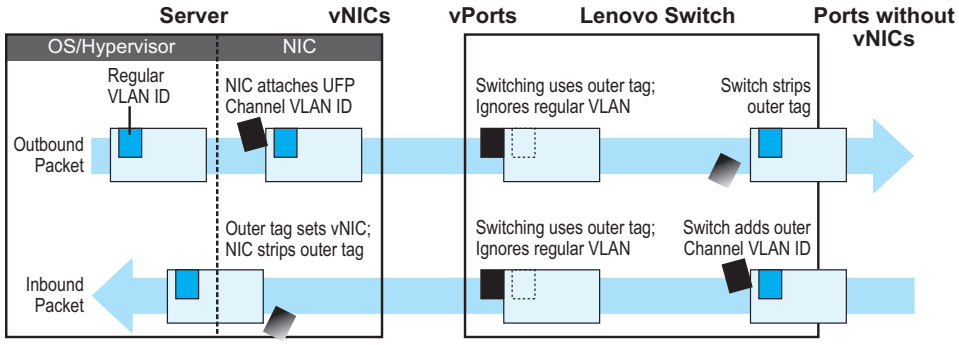
Tunnel Mode

In tunnel mode, a vPort can belong to only one VLAN. An outer tag with the vPort's VLAN ID is inserted in packets that egress the vPort. The inner VLAN tag remains unchanged. The switch processes packets based on the outer tag. When all the ports or vPorts that belong to a particular VLAN are placed in tunnel mode, they belong to one pass-through domain.

Use tunnel mode to send all VM data traffic to an upstream switch, for Layer 2 or Layer 3 processing, in one domain. In such cases, the UFP port or vPort must be in tunnel mode and the upstream switch port must be in 802.1Q trunk mode.

Note: Two vPorts on a physical port cannot be members of the same VLAN.

Figure 38. Packet pass-through in Tunnel Mode

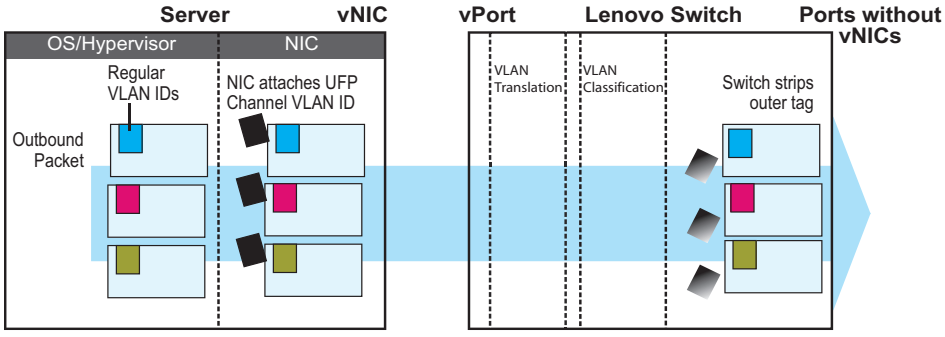


802.1Q Trunk Mode

In trunk mode, a vPort can carry packets that have inner tags that belong to up to 1024 VLANs. When UFP is enabled, the following 9 VLANs are reserved for UFP operation: 1 and 4002-4009. Each VLAN in the inner tag requires a VLAN translation entry.

Note: Two vPorts operating in trunk mode on the same physical port cannot carry the same set of VLANs in the inner tag.

Figure 39. Packet passing through in Trunk Mode



Access Mode

In access mode, a vPort carries packets with inner tags that belong to one VLAN. The vPort is associated with the VLAN defined by the command:

```
CN 4093(config_ufp_vport)# network default-vlan <2-4094>
```

Note: VLANs 4002-4009 are reserved for outer tagging.

FCoE Mode

FCoE traffic is carried by a vPort. The server-side endpoint of this virtual port will be represented through a FC vHBA. Setting a virtual port in FCoE mode will enable Priority-based Flow Control (PFC) on the physical port.

A vPort configured in FCoE mode can only be attached to a Fibre Channel (FC) VLAN. A vPort in FCoE mode operates as a local domain data path type with packets being single tagged.

Auto-VLAN Mode

When a vPort is configured in auto-VLAN mode, the vPort participates in VM discovery using VMready or 802.1Qbg. VLANs are dynamically provisioned based on VMready discovery or 802.1Qbg VM association.

When a vPort operates in auto-VLAN mode, it supports 32 VM groups. In the case of 802.1Qbg, when a vPort operates in auto-VLAN mode, the maximum number of VLANs in the inner tag is 1024 when switch is configured in standalone mode. The vPort cannot be configured in Virtual Ethernet Port Aggregator (VEPA) mode.

UFP Bandwidth Provisioning

UFP provides two modes of bandwidth provisioning for vPort: Enhanced Transmission Selection (ETS) mode and Strict Bandwidth Provisioning mode.

Enhanced Transmission Selection mode

Enhanced Transmission Selection (ETS) mode of bandwidth provisioning is useful when an end-to-end QoS framework for the entire data center, with bandwidth provisioning for different applications, is desired. ETS mode color marks traffic from point of origination to point of destination. It helps to couple QoS provisioning in the access layer with data center fabric.

Note: ETS mode requires Converged Enhanced Ethernet (CEE) to be enabled globally.

This mode functions with the ETS feature available on the CN4093. You must first define the ETS characteristics of the CN4093. Assign each vPort to the desired traffic class by assigning a system class priority. The Data Center Bridging Capabilities Exchange (DCBX) and UFP protocols propagate the configured parameters for the vPort to apply appropriate traffic coloring and shaping at the source.

When operating in this mode, traffic scheduling and bandwidth allocation behavior on switch egress is driven by the ETS class of traffic.

When two vPorts use the same traffic class configuration, the order in which switch schedules traffic at egress depends on the order the traffic arrives at egress buffer. Since bandwidth allocation is derived from traffic class rather than vNIC, switch egress doesn't differentiate between different vPort traffics.

In a virtualized environment, the hypervisor or guest VM may define its own traffic class priority. When configured this way, the priority defined by the OS or Hypervisor takes precedence over vPort-configured priority.

Use the following commands to configure ETS bandwidth provisioning:

1. Enable UFP ETS mode on a specific port:

```
CN 4093(config)# ufp port <port alias or number> qos-mode ets
```

2. Set the 802.1p priority value for the specific vPort:

```
CN 4093(config)# ufp port <port alias or number> vport <1-8>
CN 4093(config_ufp_port)# qos ets priority <0-7>
```

Note: In QoS ETS mode, only a priority of 3 is allowed for a vPort in FCoE mode.

UFP Strict Bandwidth Provisioning mode

Strict bandwidth provisioning mode configures the switch and NIC apply bidirectional bandwidth control on the vPort as per the defined configuration. By default, a bandwidth of 2.5 Gbps per vPort is guaranteed. If other vPorts are idle, the bandwidth of a vPort can be up to 10 Gbps. A minimum bandwidth of 1 Gbps is provisioned, which can be raised by 100 Mbps increments. The sum of the minimum bandwidth guaranteed for all vPorts together cannot exceed the capacity of the physical link. A vPort can also be configured with a maximum bandwidth.

This mode works with the port scheduler to avoid unintended packet drops due to policing through EFP metering block. If flow control is enabled, the switch provides a no-drop packet forwarding behavior, which improves end-to-end TCP-throughput performance.

Note: If a vPort is configured with low upper limit, it might lead to head-of-line congestion on the egress port.

ETS mode is disabled when strict bandwidth provisioning mode is enabled. By default, uplink ports have a separate traffic class for storage traffic with guaranteed bandwidth. The rest of the bandwidth is shared equally among other traffic.

Use the following commands to configure strict bandwidth provisioning:

1. Enable UFP Strict Bandwidth Provisioning mode on a specific port:

```
CN 4093(config)# ufp port <port alias or number> qos-mode bw
```

2. Configure the bandwidth settings for a specific vPort:

```
CN 4093(config)# ufp port <port alias or number> vport <1-8>
CN 4093(config_ufp_vport)# qos bandwidth {max|min} <10-100>

min - Sets minimum guaranteed bandwidth
max - Sets maximum allowed bandwidth
```

Note: Total minimum guaranteed bandwidth of enabled vPorts on a physical switch port needs to be 100%.

Using UFP with Other CN4093 10 Gb Converged Scalable Switch Features

UFP works with other CN4093 features, as described with limitations and details.

Layer 2 Failover

UFP failover can be configured with auto-monitoring or manual monitoring. In auto-monitoring, a vPort is automatically associated with a Failover trigger if it has any VLAN in common with the monitor ports.

Layer 2 failover is not supported on UFP ports in auto mode.

For more information on failover, see "Layer 2 Failover" on page 519.

For an example configuration, see "Example 8: Layer 2 Failover Configuration" on page 383.

Increased VLAN Limits

Configured with UFP and VLANs, a vPort can support up to `1024` VLANs. A UFP port supports `1024` VLANs on the switch in standalone mode.

For more information on VLAN configuration, see "VLANs" on page 141.

Private VLANs

It supports the following Private VLAN modes in UFP vPorts:

- Disabled
- Trunk
- Promiscuous
- Host

The following are the criteria of these Private VLAN modes:

- Private-VLAN mode is `disabled`:
  - Allows only non-private domain.
- Private-VLAN mode is `trunk`:
  - Allows both primary and secondary VLAN which belong to the private VLAN domain.
  - Allows non-private VLAN domains.
  - vPorts belonging to the same UFP port cannot be in the same private-VLAN domain.
  - For the traffic of a specific VLAN to be passed through a port, requires the port to be explicitly added to that VLAN. For example, for a UFP port to pass traffic in private VLAN primary 100 and secondary 10, the port must be added to both VLAN 100 and VLAN 10.
  - Since this port type is intended to be functional as an ISL port, the isolate-VLAN is allowed to pass traffic through this port type.

- Private-VLAN mode is `promiscuous`:
  - Allows only primary VLAN.
  - There can be multiple Private VLAN domains.
  - The Private VLAN domains must be unique for vPorts belonging to the same UFP port.
- Private-VLAN mode is `host`:
  - Allows only ONE secondary VLAN. In the case of a vPort is in network trunk mode, there will be multiple VLANs assigned to the vPort, but there will still be only ONE secondary VLAN. The other VLANs will be in a different private VLAN domain.
  - Warns if no primary VLAN is associated with the secondary VLAN assigned to a vPort.

UFP with private VLANs is supported under the following limitations:

- vPorts from the same physical port cannot belong to the same private VLAN domain.
- vPorts cannot be configured with a primary VLAN as a default VLAN, only with secondary VLANs.
- UFP ports cannot have switchport mode private-VLAN enabled on them.
- Private VLAN is supported only on vPorts configured with trunk or access mode.
- UFP cannot be configured on promiscuous ports.

For more information on private VLANs, see "Private VLANs" on page 157 .

VMReady

Configuring with UFP and VMReady, the CN4093 can support up to 32 VMGroups with UFP vPorts in auto mode.

VMReady is supported only on a vPort which is configured in auto-VLAN mode.

For more information on VMReady, see "VMready" on page 283.

802.1Qbg

Configured with Edge Virtual Bridging (EVB), UFP supports up to `1024` VLANs on a vPort. In Stacking mode, UFP supports up to 256 VLANs on a single vPort.

EVB is supported only on a vPort which is configured in auto-VLAN mode.

For more information on EVB, see "Edge Virtual Bridging" on page 353.

UFP Configuration Examples

Following is an example configuration of UFP vPorts in access mode.

Example 1: Access Mode

1. Turn on UFP.

```
CN4093(config)# ufp enable
```

2. Configure internal port as UFP.

```
CN4093(config)# ufp port INTA1 enable

Warning: "Tagging/Trunk-mode" is enabled on UFP port INTA1
```

3. Configure virtual port.

```
CN4093(config)# ufp port INTA1 vport 1
```

4. Configure vPort access mode.

```
CN4093(config_ufp_vport)# network mode access
```

5. Configure vPort default VLAN.

```
CN4093(config_ufp_vport)# network default-vlan 100
```

6. Specify QoS parameters for the vPort.

```
CN4093(config_ufp_vport)# qos bandwidth min 30          (in percentage)
CN4093(config_ufp_vport)# qos bandwidth max 90          (in percentage)
```

7. Enable vPort.

```
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit
```

8. Configure PVID/Native VLAN for external port 1.

```
CN4093(config)# interface port EXT1
CN4093(config-if)# switchport mode access
CN4093(config-if)# switchport access vlan 100
```

Example 2: Trunk Mode

Following is an example configuration of UFP vPorts in trunk mode.

1. Turn on UFP.

```
CN 4093(config)# ufp enable
```

2. Configure internal port 1 as UFP.

```
CN4093(config)# ufp port INTA1 enable

Warning: "Tagging/Trunk-mode" is enabled on UFP port INTA1
```

3. Configure virtual port.

```
CN4093(config)# ufp port INTA1 vport 1
```

4. Configure vPort trunk mode.

```
CN4093(config_ufp_vport)# network mode trunk
```

5. Configure vPort default VLAN.

```
CN4093(config_ufp_vport)# network default-vlan 100
```

6. Specify QoS parameters for the vPort.

```
CN4093(config_ufp_vport)# qos bandwidth min 15                (in percentage)
CN4093(config_ufp_vport)# qos bandwidth max 80                (in percentage)
```

7. Enable vPort.

```
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit
```

8. Configure internal port 2 as UFP.

```
CN4093(config)# ufp port INTA2 enable

Warning: "Tagging/Trunk-mode" is enabled on UFP port INTA2
```

9. Configure virtual port.

```
CN4093(config)# ufp port INTA2 vport 3
```

10. Configure vPort trunk mode.

```
CN4093(config_ufp_vport)# network mode trunk
```

11. Configure the vPort default VLAN.

```
CN4093(config_ufp_vport)# network default-vlan 100
```

12. Optionally, you can disable tagging on the vPort.

```
CN4093(config_ufp_vport)# no network default-tag
```

13. Specify QoS parameters for the vPort.

```
CN4093(config_ufp_vport)# qos bandwidth min 15          (in percentage)
CN4093(config_ufp_vport)# qos bandwidth max 95          (in percentage)
```

14. Enable vPort.

```
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit
```

15. Enable tagging/trunk mode on external port 1.

```
CN4093(config)# interface port EXT1
CN4093(config-if)# switchport mode trunk
CN4093(config-if)# switchport trunk native vlan 100
CN4093(config-if)# switchport trunk allowed vlan add 200,300
CN4093(config-if)# exit
```

16. Configure VLAN 200 parameters.

```
CN4093(config)# vlan 200
CN4093(config-vlan)# vmember INTA1.1
CN4093(config-vlan)# vmember INTA2.3
CN4093(config-vlan)# exit
```

17. Configure VLAN 300 parameters.

```
CN4093(config)# vlan 300
CN4093(config-vlan)# vmember INTA1.1
CN4093(config-vlan)# vmember INTA2.3
CN4093(config-vlan)# exit
```

Example 3: Auto-VLAN Mode with VMready

1. Turn on UFP.

```
CN 4093(config)# ufp enable
```

2. Configure internal port 1 as UFP.

```
CN4093(config)# ufp port INTA1 enable

Warning: "Tagging/Trunk-mode" is enabled on UFP port INTA1
```

3. Configure virtual port.

```
CN4093(config)# ufp port INTA1 vport 1
```

4. Configure vPort default VLAN.

```
CN4093(config_ufp_vport)# network default-vlan 100
```

5. Configure vPort auto mode.

```
CN4093(config_ufp_vport)# network mode auto
```

Note: VLAN is dynamically added by VMready.

6. Specify QoS parameters for the vPort.

```
CN4093(config_ufp_vport)# qos bandwidth min 20          (in percentage)
CN4093(config_ufp_vport)# qos bandwidth max 90          (in percentage)
```

7. Enable vPort.

```
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit
```

8. Enable virtual machine groups.

```
CN4093(config)# virt enable
```

9. Configure the VMware settings.

```
CN4093(config)# virt vmware vcspec 10.100.14.195 Administrator noauth
7cee7fa528e02aa036b6b6e6eb508952cdaed2acb702182cf62208fa72dec13fb19d6fec2
fac6598d19b
8f45acff3f6a1e237ae3c984709f874f61aecd2ede7a
```

10. Create a distributed VMGroup..

```
CN4093(config)# virt vmprofile “vlan 30”
CN4093(config)# virt vmprofile edit “vlan 30” vlan 30
CN4093(config)# virt vmgroup 1 profile “vlan30”
```

11. Verify the virtual machine settings.

```
CN4093(config)# show virt vm
```

12. Add the virtual machine associated with the vPort to the VMGroup.

```
CN4093(config)# virt vmgroup 1 vm 1
```

13. Verify the VMGroup associations.

```
CN4093(config)# show virt vm
```

Example 4: Auto-VLAN Mode with Edge Virtual Bridging

Following is an example configuration of UFP vPorts in auto mode.

1. Turn on UFP.

```
CN 4093(config)# ufp enable
```

2. Configure internal port 1 as UFP.

```
CN4093(config)# ufp port INTA1 enable
Warning: "Tagging/Trunk-mode" is enabled on UFP port INTA1
```

3. Configure a vPort.

```
CN4093(config_ufp_vport)# ufp port INTA1 vport 1
```

4. Configure the vPort's default VLAN.

```
CN4093(config_ufp_vport)# network default-vlan 20
```

5. Set the vPort to tunnel mode.

```
CN4093(config_ufp_vport)# network mode auto
```

Note: VLAN is dynamically added by 802.1Qbg.

6. Configure the EVG profile for the vPort.

```
CN4093(config_ufp_vport)# evb profile 1
```

7. Specify QoS parameters for the vPort.

```
CN4093(config_ufp_vport)# qos bandwidth min 10          (in percentage)
CN4093(config_ufp_vport)# qos bandwidth max 80          (in percentage)
```

8. Enable vPort.

```
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit
```

9. Configure the Edge Virtual Bridging profile.

```
CN4093(config)# virt evb profile 1
CN4093(conf-evbprof)# reflective-relay
CN4093(conf-evbprof)# vsi-discovery
CN4093(conf-evbprof)# exit
```

10. Configure the Edge Virtual Bridging database.

```
CN4093(config)# virt evb vsidb 1
CN4093(conf-vsidb)# host 10.100.48.20
CN4093(conf-vsidb)# filepath vsitypes
CN4093(conf-vsidb)# exit
```

Note: VLANs in the database are dynamically added by 802.1Qbg.

Example 5: Tunnel Mode

Following is an example configuration of UFP vPorts in tunnel mode.

1. Turn on UFP.

```
CN 4093(config)# ufp enable
```

2. Configure internal port as UFP.

```
CN4093(config)# ufp port INTA1 enable

Warning: "Tagging/Trunk-mode" is enabled on UFP port INTA1
```

3. Configure virtual port.

```
CN4093(config)# ufp port INTA1 vport 1
```

4. Configure vPort tunnel mode.

```
CN4093(config_ufp_vport)# network mode tunnel
```

5. Configure vPort default VLAN.

```
CN4093(config_ufp_vport)# network default-vlan 4000
```

6. Specify the QoS parameters for the vPort.

```
CN4093(config_ufp_vport)# qos bandwidth min 15          (in percentage)
CN4093(config_ufp_vport)# qos bandwidth max 95          (in percentage)
```

7. Enable the vPort.

```
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit
```

8. Configure tagging of ingress frames with the port's VLAN ID on external port 1.

```
CN4093(config)# interface port EXT1
CN4093(config-if)# tagpvid-ingress
CN4093(config-if)# no vlan dot1q tag native
CN4093(config-if)# switchport access vlan 4000
CN4093(config-if)# exit
```

Example 6: FCoE Mode

Following is an example configuration of UFP vPorts in FCoE mode.

This example is consistent with the network shown in Figure 32 on page 304.

1. Enable CEE.

```
CN4093(config)# cee enable
```

2. Enable FIPs.

```
CN4093(config)# fcoe fips enable
```

3. Turn on UFP.

```
CN 4093(config)# ufp enable
```

4. Configure internal port as UFP.

```
CN4093(config)# ufp port INTA1 enable

Warning: "Tagging/Trunk-mode" is enabled on UFP port INTA1
```

5. Configure virtual port.

```
CN4093(config)# ufp port INTA1 vport 2
```

6. Configure vPort FCoE mode.

```
CN4093(config_ufp_vport)# network mode fcoe
```

7. Configure vPort default VLAN.

```
CN4093(config_ufp_vport)# network default-vlan 1002
```

8. Specify QoS parameters for the vPort.

```
CN4093(config_ufp_vport)# qos bandwidth min 20                    (in percentage)
CN4093(config_ufp_vport)# qos bandwidth max 85                    (in percentage)
```

9. Enable vPort.

```
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit
```

10. Enable tagging/trunk mode on external port.

```
CN4093(config)# interface port EXT4
CN4093(config-if)# switchport mode trunk
CN4093(config-if)# switchport trunk native vlan 1
CN4093(config-if)# switchport trunk allowed vlan add 1,1002
CN4093(config-if)# exit
```

Example 7: Private VLAN Configuration

Follow this procedure to configure a Private VLAN.

1. Select a VLAN and define the Private VLAN type as primary.

```
CN 4093(config)# vlan 700
CN 4093(config)# private-vlan primary
CN 4093(config)# exit
```

2. Configure a promiscuous port for VLAN 700.

```
CN4093(config)# interface port INTA10
CN4093(config-if)# switchport mode private-vlan
CN4093(config-if)# switchport private-vlan mapping 700
CN4093(config-if)# exit
```

3. Configure three secondary VLANs: isolated VLAN and community VLAN.

```
CN4093(config)# vlan 701
CN4093(config-vlan)# private-vlan isolated
CN4093(config-vlan)# exit
CN4093(config)# vlan 702
CN4093(config-vlan)# private-vlan community
CN4093(config-vlan)# exit
CN4093(config)# vlan 703
CN4093(config-vlan)# private-vlan community
CN4093(config-vlan)# exit
```

4. Map secondary VLANs to primary VLAN.

```
CN4093(config)# vlan 700-703
CN4093(config-vlan)# stg 1
CN4093(config-vlan)# exit
CN4093(config)# vlan 700
CN4093(config-vlan)# private-vlan association 701,702,703
CN4093(config-vlan)# exit
```

5. Set up vPorts on ports 1 and 2.

```
CN4093(config)# ufp port INTA1 enable
CN4093(config)# ufp port INTA1 vport 1
CN4093(config-ufp-vport)# network private-vlan trunk
CN4093(config-ufp-vport)# network default-vlan 100
CN4093(config-ufp-vport)# network mode trunk
CN4093(config-ufp-vport)# enable
CN4093(config-ufp-vport)# exit

CN4093(config)# ufp port INTA2 enable
CN4093(config)# ufp port INTA2 vport 1
CN4093(config-ufp-vport)# network private-vlan promiscuous
CN4093(config-ufp-vport)# network default-vlan 200
CN4093(config-ufp-vport)# network mode trunk
CN4093(config-ufp-vport)# enable
CN4093(config-ufp-vport)# exit

CN4093(config)# ufp port INTA3 enable
CN4093(config)# ufp port INTA3 vport 1
CN4093(config-ufp-vport)# network private-vlan host
CN4093(config-ufp-vport)# network default-vlan 300
CN4093(config-ufp-vport)# network mode trunk
CN4093(config-ufp-vport)# enable
CN4093(config-ufp-vport)# exit

CN4093(config)# vlan 700
CN4093(config-vlan)# vmember INTA1.1
CN4093(config-vlan)# vmember INTA2.1
CN4093(config-vlan)# exit
CN4093(config)# vlan 701
CN4093(config-vlan)# vmember INTA1.1
CN4093(config-vlan)# exit
CN4093(config)# vlan 702
CN4093(config-vlan)# vmember INTA1.1
CN4093(config-vlan)# exit
CN4093(config)# vlan 703
CN4093(config-vlan)# vmember INTA1.1
CN4093(config-vlan)# vmember INTA3.1
CN4093(config-vlan)# exit
```

6. Verify the configuration.

```
CN4093(config)# show vlan private-vlan
```

Example 8: Layer 2 Failover Configuration

While configuring a failover trigger, you cannot use the `member` command for a physical port that has vPorts configured. Instead, you must use the `vmember` command to add the vPorts as members of a failover trigger. The following example includes the commands to configure a failover trigger using a physical port INTA8 (UFP not enabled) and vPorts INTA9.1, INTA10.2 and INTA11.3 configured on UFP-enabled physical ports INTA9, INTA10 and INTA11.

See "Example 1: Access Mode" on page 374 for steps to configure a vPort in access mode. Follow the steps below for configuring the failover trigger:

1. Enable failover globally:

```
CN4093(config)# failover enable
```

2. Configure trigger 1 and add monitor and control ports:

```
CN4093(config)# failover trigger 1 mmon monitor member EXT1
CN4093(config)# failover trigger 1 mmon control member INTA8
CN4093(config)# failover trigger 1 mmon control vmember INTA9.1,
INTA10.2,INTA11.3
```

Note: If you try to add a physical port (that has vPorts configured) as a member of a trigger, you may see the following error message when you enable the trigger:

```
CN4093(config)# failover trigger 1 ena
Failover Error: trigger 1 physical port INTA9 has virtual ports.
```

3. Enable failover trigger:

```
CN4093(config)# failover trigger 1 enable
```

Example 9: 8 vPorts with ETS bandwidth provisioning mode

Follow this procedure to configure 8 vPorts for a single UFP port with ETS bandwidth provisioning mode.

1. Configure each individual vPort of a specific port:

```
CN4093(config)# ufp port INTA10 vport 1
CN4093(config_ufp_vport)# network mode access
CN4093(config_ufp_vport)# network default-vlan 101
CN4093(config_ufp_vport)# qos ets priority 0
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit

CN4093(config)# ufp port INTA10 vport 2
CN4093(config_ufp_vport)# network mode fcoe
CN4093(config_ufp_vport)# network default-vlan 1002
CN4093(config_ufp_vport)# qos ets priority 3
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit

CN4093(config)# ufp port INTA10 vport 3
CN4093(config_ufp_vport)# network mode trunk
CN4093(config_ufp_vport)# network default-vlan 21
CN4093(config_ufp_vport)# qos ets priority 1
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit

CN4093(config)# ufp port INTA10 vport 4
CN4093(config_ufp_vport)# network mode trunk
CN4093(config_ufp_vport)# network default-vlan 400
CN4093(config_ufp_vport)# qos ets priority 2
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit

CN4093(config)# ufp port INTA10 vport 5
CN4093(config_ufp_vport)# network mode trunk
CN4093(config_ufp_vport)# network default-vlan 43
CN4093(config_ufp_vport)# qos ets priority 4
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit

CN4093(config)# ufp port INTA10 vport 6
CN4093(config_ufp_vport)# network mode tunnel
CN4093(config_ufp_vport)# network default-vlan 65
CN4093(config_ufp_vport)# qos ets priority 5
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit

CN4093(config)# ufp port INTA10 vport 7
CN4093(config_ufp_vport)# network mode tunnel
CN4093(config_ufp_vport)# network default-vlan 98
CN4093(config_ufp_vport)# qos ets priority 6
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit

CN4093(config)# ufp port INTA10 vport 8
CN4093(config_ufp_vport)# network mode tunnel
CN4093(config_ufp_vport)# network default-vlan 654
CN4093(config_ufp_vport)# qos ets priority 7
CN4093(config_ufp_vport)# enable
CN4093(config_ufp_vport)# exit
```

2. Configure ETS mode as the UFP QoS mode for port INTA10:

```
CN4093(config)# ufp port INTA10 qos-mode ets
```

3. Enable UFP on port INTA10:

```
CN4093(config)# ufp port INTA10 enable
```

4. Globally enable Converged Enhanced Ethernet (CEE):

```
CN4093(config)# cee enable
```

5. Globally enable UFP:

```
CN4093(config)# ufp enable
```

Chapter 22. Switch Partition

Switch Partition (SPAR) enables consolidation of multiple network partitions within an embedded switch. SPARs divide the data plane of a physical switch into independent switching domains. Switch partitions are isolated from each other. Traffic originating in one SPAR stays local to that SPAR. Within a partitioned switch, traffic from one SPAR is never delivered to another SPAR. Traffic from one SPAR can, however, be delivered to another SPAR by traversing an upstream link and switch.

Each individual SPAR requires exactly one uplink, which can be a port, a port channel, or an LACP group. Limiting SPAR connectivity to one external uplink prevents the creation of loops.

SPAR operates as a Layer 2 broadcast network. Hosts on the same VLAN, attached to a SPAR, can communicate with each other and with the upstream switch. Hosts on the same VLAN, but attached to different SPARs, communicate via the upstream switch.

SPAR Processing Modes

SPAR operates in two processing modes. The default mode is pass-through domain.

- Local Domain: In local-domain processing mode, VLAN classification and assignment is based on the user-defined VLAN.
- Pass-through Domain: In pass-through domain processing mode, VLAN classification and assignment is based on the outer tag, which contains the unique domain VLAN ID of the SPAR. The inner tag with the user-defined VLAN remains unchanged.

Local Domain Processing

Each SPAR on a switch has a unique VLAN ID, which separates data between SPARs. If multiple networks share the uplink, the upstream switch port must be configured as a 802.1Q trunk port so it can process multiple VLAN traffic from a SPAR. The SPAR domain uses a single uplink port or LAG shared among all the VLANs. For link redundancy or greater bandwidth, the uplinks can be grouped as static or LACP LAG.

If a VLAN is defined on multiple SPARs, the egress port mask is used to prevent communication between the SPARs in the same local domain VLAN. Since port membership of each SPAR is unique, the egress port mask ensures that different SPAR ports in the same local domain VLAN do not communicate with each other.

In local domain processing, all SPAR ports must have the following settings:

- Tagging/Trunk mode must be enabled.
- Ingress VLAN tagging is disabled on all SPAR ports.
- PVID/Native VLAN is based on any VLAN defined in SPAR.

```
CN 4093(config)# interface port <num>
CN 4093(config-if)# switchport trunk native vlan <VLAN number>
```

Pass-Through Domain Processing

Pass-through domain processing is the default operating mode for SPAR when performing L2 switching based on an outer tag.

In pass-through processing mode, each SPAR is identified by its unique VLAN domain ID. Packets are classified based on the SPAR domain ID (outer tag). SPAR ports must be configured in tunnel mode.

SPAR provides single or multiple VLAN connectivity through a single uplink port or LAG (static or LACP). VLAN definition within the SPAR domain is not required.

Pass-through domain operates in Q-In-Q mode. Inside SPAR, different user-defined VLAN traffic is classified into single S-VLAN (service VLAN) associated with the SPAR.

Although the uplink can be shared by multiple networks using the pass-through domain, SPAR will not be server-VLAN aware. Hence, multiple VLAN traffic will be mixed together in a single broadcast domain, that is, broadcast traffic on different VLANs from the upstream network will reach all servers attached to the SPAR pass-through domain. The servers drop the packets if they do not belong to the desired VLAN. The pass-through implementation uses ingress VLAN tagging, that is, tagpvid-ingress is enabled on all SPAR ports.

In pass-through domain processing mode, all SPAR ports must have the following settings:

- PVID/Native VLAN tagging is disabled.
- Ingress VLAN tagging is enabled on all SPAR ports.
- PVID/Native VLAN is based on the SPAR DVLAN.

```
CN 4093(config)# interface port <num>
CN 4093(config-if)# switchport trunk native vlan <VLAN number>
```

Limitations

The following limitations apply:

- UFP and SPAR cannot be configured together.
- LAGs must first be configured for SPAR before they can be used. Static or Link Aggregation Control Protocol (LACP) LAGs created on the global switch cannot reference any SPAR ports. Use the commands in the following menus to define LAGs in the SPAR context:

```
CN 4093(config)# spar <num>
CN 4093(config-spar)# uplink ?

  adminkey     Set lacp trunk for uplink
  port         Set external port for uplink
  PortChannel  Set portchannel for uplink
```

```
CN 4093(config)# portchannel ?

 <1-64>    PortChannel group
 <65-128>  LACP PortChannel group
 thash     Port Channel hash configuration
```

- ACLs defined on the global switch can be used for SPAR ports. However, the following restrictions apply:
  - An ACL cannot be shared across SPAR ports if:
    - An exit port (`CN 4093(config)# access-control list` *<number>* `egress-port port` *<number>*) is used as a filtering criteria and the exit port does not belong to the same SPAR as the port on which the ACL is applied.
    - A monitor port is used as a filtering criteria, and the monitor port does not belong to the same SPAR as the mirrored port and is not defined on the global switch.
  - These ACL restrictions apply to all ACLs defined in an ACL group.
- Port mirroring can be configured on SPAR ports, but the monitor port must either belong to the same SPAR as the mirrored port or must be defined on the global switch.
- Layer 2 failover features can be configured on SPAR ports. However, the Layer 2 failover Auto Monitor (AMON) option is not supported. Only the Layer 2 failover Manual Monitor (MMON) option can be used when all ports defined within the trigger belong to the same SPAR.

Unsupported Features

The following features are not supported when SPAR is configured:

- 802.1x
- Edge Virtual Bridging
- Fibre Channel over Ethernet (FCoE)
- Hotlinks
- IGMP
- Layer 3 Configuration
- Management VLAN
- Private VLAN
- Protocol VLAN
- sFlow
- Stacking
- STP, RSTP, MRSTP, PVST
- UFP
- vLAG
- VMAP
- VMready
- VNIC

SPAR VLAN Management

SPAR VLANs use the same 4000 VLAN space available for other applications/features on the switch. The VLAN ID can be in the range of 2 - 4094. VLAN 1 and the management VLAN 4095 are reserved for the global switch context.

A VLAN assigned to a SPAR cannot be used for any other switch application. Similarly, VLAN used by any other switch application cannot be assigned to a SPAR.

SPAR member ports cannot be members of any other VLAN.

Example Configurations

The following are examples of SPAR pass through and local domain configurations.

Pass Through Configuration

This example describes configuration of SPAR 1 in pass-through mode with internal server ports INTA5 through INTA10, with a single port, EXT1.

1. Create SPAR 1.

```
CN 4093(config)# spar 1
```

   Each SPAR is identified with a number that ranges 1 though 8.

2. Add a single uplink port to SPAR 1.

```
CN 4093(config-spar)# uplink port EXT1
```

3. Set the mode of the SPAR to passthrough

```
CN 4093(config-spar)# domain mode passthrough
```

4. Configure SPAR VLAN to 4081.

```
CN 4093(config-spar)# domain default vlan 4081
```

5. Add ports INTA5 through INTA10 to SPAR 1.

```
CN 4093(config-spar)# domain default member INTA5-INTA10
```

6. Enable SPAR 1.

```
CN 4093(config-spar)# enable
```

Local Domain Configuration

This example demonstrates how to create a SPAR in local-domain mode consisting of internal server ports INTA11-INTA14 and a single uplink port, EXT 2.

1. Create SPAR 2.

```
CN 4093(config)# spar 2
```

2. Add uplink port EXT 2 to SPAR 2.

```
CN 4093(config-spar)# uplink port EXT2
```

3. Set the SPAR to local domain mode.

```
CN 4093(config-spar)# domain mode local
```

4. Configure SPAR VLAN to 4082.

```
CN 4093(config-spar)# domain default vlan 4082
```

5. Add server ports INTA11 through INTA14.

```
CN 4093(config-spar)# domain default member INTA11-INTA14
```

6. Configure the VLANs for SPAR 2.
   Each SPAR has a set of local domains numbered 1 through 32, each of which identifies an allowed VLAN.

   The following steps create three local domains: VLAN, 10, 20, and 30

7. Create local domain 1, assign VLAN 10, and specify the SPAR ports that are members of the that VLAN.

```
CN 4093(config-spar)# domain local 1 vlan 10
CN 4093(config-spar)# domain local 1 member INTA11-INTA14
CN 4093(config-spar)# domain local 1 enable
```

8. Create local domain 2, assign VLAN 20, and specify the SPAR ports that are members of the that VLAN.

```
CN 4093(config-spar)# domain local 2 vlan 20
CN 4093(config-spar)# domain local 2 member INTA11-INTA14
CN 4093(config-spar)# domain local 2 enable
```

9. Create local domain 3, assign VLAN 30, and specify the SPAR ports that are members of the that VLAN.

```
CN 4093(config-spar)# domain local 3 vlan 30
CN 4093(config-spar)# domain local 3 member INTA11-INTA14
CN 4093(config-spar)# domain local 3 enable
```

10. Enable SPAR 2.

```
CN 4093(config-spar)# enable
```

Part 5: IP Routing

This section discusses Layer 3 switching functions. In addition to switching traffic at near line rates, the application switch can perform multi-protocol routing. This section discusses basic routing and advanced routing protocols:

- Basic Routing
- Routing Information Protocol (RIP)
- Internet Group Management Protocol (IGMP)
- Border Gateway Protocol (BGP)
- Open Shortest Path First (OSPF)

Chapter 23. Basic IP Routing

This chapter provides configuration background and examples for using the CN4093 10 Gb Converged Scalable Switch (CN4093) to perform IP routing functions. The following topics are addressed in this chapter:

IP Routing Benefits

The CN4093 uses a combination of configurable IP switch interfaces and IP routing options. The switch IP routing capabilities provide the following benefits:

- Connects the server IP subnets to the rest of the backbone network.
- Provides the ability to route IP traffic between multiple Virtual Local Area Networks (VLANs) configured on the switch.

Routing Between IP Subnets

The physical layout of most corporate networks has evolved over time. Classic hub/router topologies have given way to faster switched topologies, particularly now that switches are increasingly intelligent. The CN4093 is intelligent and fast enough to perform routing functions on par with wire-speed Layer 2 switching.

The combination of faster routing and switching in a single device provides another service—it allows you to build versatile topologies that account for legacy configurations.

Consider an example in which a corporate campus has migrated from a router-centric topology to a faster, more powerful, switch-based topology. As is often the case, the legacy of network growth and redesign has left the system with a mix of illogically distributed subnets.

This is a situation that switching alone cannot cure. Instead, the router is flooded with cross-subnet communication. This compromises efficiency in two ways:

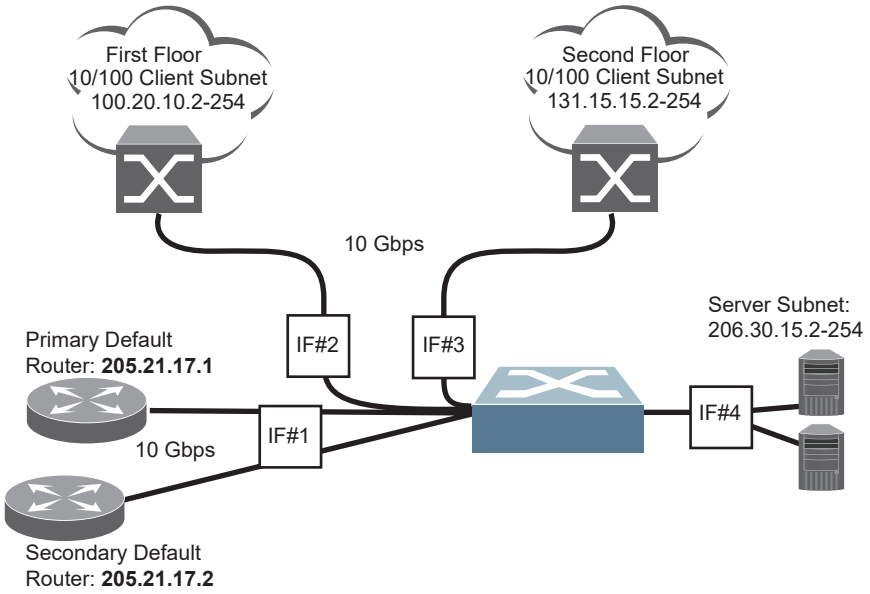
- Routers can be slower than switches. The cross-subnet side trip from the switch to the router and back again adds two hops for the data, slowing throughput considerably.
- Traffic to the router increases, increasing congestion.

Even if every end-station could be moved to better logical subnets (a daunting task), competition for access to common server pools on different subnets still burdens the routers.

This problem is solved by using CN4093s with built-in IP routing capabilities. Cross-subnet LAN traffic can now be routed within the switches with wire speed Layer 2 switching performance. This not only eases the load on the router but saves the network administrators from reconfiguring each and every end-station with new IP addresses.

Take a closer look at the CN4093 in the following configuration example:

Figure 40. Switch-Based Routing Topology



The CN4093 connects the Gigabit Ethernet and Fast Ethernet LAGs from various switched subnets throughout one building. Common servers are placed on another subnet attached to the switch. A primary and backup router are attached to the switch on yet another subnet.

Without Layer 3 IP routing on the switch, cross-subnet communication is relayed to the default gateway (in this case, the router) for the next level of routing intelligence. The router fills in the necessary address information and sends the data back to the switch, which then relays the packet to the proper destination subnet using Layer 2 switching.

With Layer 3 IP routing in place on the CN4093, routing between different IP subnets can be accomplished entirely within the switch. This leaves the routers free to handle inbound and outbound traffic for this group of subnets.

Subnet Routing Example

Prior to configuring, you must be connected to the switch Command Line Interface (CLI) as the administrator.

Note: For details about accessing and using any of the menu commands described in this example, see the *Enterprise NOS Command Reference*.

1. Assign an IP address (or document the existing one) for each router and client workstation.

   In the example topology in Figure 40 on page 398, the following IP addresses are used:

   Table 33. *Subnet Routing Example: IP Address Assignments*

Subnet	Devices	IP Addresses
1	Primary and Secondary Default Routers	205.21.17.1 and 205.21.17.2
2	First Floor Client Workstations	100.20.10.2-254
3	Second Floor Client Workstations	131.15.15.2-254
4	Common Servers	206.30.15.2-254

2. Assign an IP interface for each subnet attached to the switch.

   Since there are four IP subnets connected to the switch, four IP interfaces are needed:

   Table 34. *Subnet Routing Example: IP Interface Assignments*

Interface	Devices	IP Interface Address
IF 1	Primary and Secondary Default Routers	205.21.17.3
IF 2	First Floor Client Workstations	100.20.10.1
IF 3	Second Floor Client Workstations	131.15.15.1
IF 4	Common Servers	206.30.15.1

   IP interfaces are configured using the following commands:

```
CN 4093(config)# interface ip 1                (Select IP interface 1)
CN 4093(config-ip-if)# ip address 205.21.17.3 255.255.255.0 enable
CN 4093(config-vlan)# exit
CN 4093(config)# interface ip 2                (Select IP interface 2)
CN 4093(config-ip-if)# ip address 100.20.10.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 3                (Select IP interface 3)
CN 4093(config-ip-if)# ip address 131.15.15.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 4                (Select IP interface 4)
CN 4093(config-ip-if)# ip address 206.30.15.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
```

3. Set each server and workstation's default gateway to the appropriate switch IP interface (the one in the same subnet as the server or workstation).

4. Configure the default gateways to the routers' addresses.

   Configuring the default gateways allows the switch to send outbound traffic to the routers:

```
CN 4093(config)# ip gateway 1 address 205.21.17.1 enable

CN 4093(config)# ip gateway 2 address 205.21.17.2 enable
```

5. Verify the configuration.

```
CN 4093(config)# show interface ip
```

   Examine the resulting information. If any settings are incorrect, make the appropriate changes.

Using VLANs to Segregate Broadcast Domains

In the previous example, devices that share a common IP network are all in the same broadcast domain. If you want to limit the broadcasts on your network, you could use VLANs to create distinct broadcast domains. For example, as shown in the following procedure, you could create one VLAN for the client trunks, one for the routers, and one for the servers.

In this example, you are adding to the previous configuration.

1. Determine which switch ports and IP interfaces belong to which VLANs.

   The following table adds port and VLAN information:

   Table 35. *Subnet Routing Example: Optional VLAN Ports*

VLAN	Devices	IP Interface	Switch Port	VLAN #
1	First Floor Client Workstations	2	EXT1	1
	Second Floor Client Workstations	3	EXT2	1
2	Primary Default Router	1	EXT3	2
	Secondary Default Router	1	EXT4	2
3	Common Servers 1	4	INT5A	3
	Common Servers 2	4	INT6A	3

2. Add the switch ports to their respective VLANs.

   The VLANs shown in Table 35 are configured as follows:

```
CN 4093(config)# vlan 1
CN 4093(config-vlan)# exit
CN 4093(config)# interface port ext1,ext2          (Add ports to VLAN 1)
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan add 1
CN 4093(config-if)# exit

CN 4093(config)# vlan 2
CN 4093(config-vlan)# exit
CN 4093(config)# interface port ext3,ext4          (Add ports to VLAN 2)
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan add 2
CN 4093(config-if)# exit

CN 4093(config)# vlan 3
CN 4093(config-vlan)# exit
CN 4093(config)# interface port inet5a,int6a       (Add ports to VLAN 3)
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan add 3
CN 4093(config-if)# exit
```

Each time you add a port to a VLAN, you may get the following prompt:

```
Port 4 is an untagged port and its current PVID is 1.
Confirm changing PVID from 1 to 2 [y/n]?
```

Enter **y** to set the default Port VLAN ID (PVID) for the port.

3. Add each IP interface to the appropriate VLAN.

   Now that the ports are separated into three VLANs, the IP interface for each subnet must be placed in the appropriate VLAN. From Table 35, the settings are made as follows:

```
CN 4093(config)# interface ip 1                 (Select IP interface 1)
CN 4093(config-ip-if)# vlan 2                   (Add VLAN 2)
CN 4093(config-vlan)# exit
CN 4093(config)# interface ip 2                 (Select IP interface 2)
CN 4093(config-ip-if)# vlan 1                   (Add VLAN 1)
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 3                 (Select IP interface 3)
CN 4093(config-ip-if)# vlan 1                   (Add VLAN 1)
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 4                 (Select IP interface 4)
CN 4093(config-ip-if)# vlan 3                   (Add VLAN 3)
CN 4093(config-ip-if)# exit
```

4. Verify the configuration.

```
CN 4093(config)# show vlan
CN 4093(config)# show interface information
CN 4093(config)# show interface ip
```

   Examine the resulting information. If any settings are incorrect, make the appropriate changes.

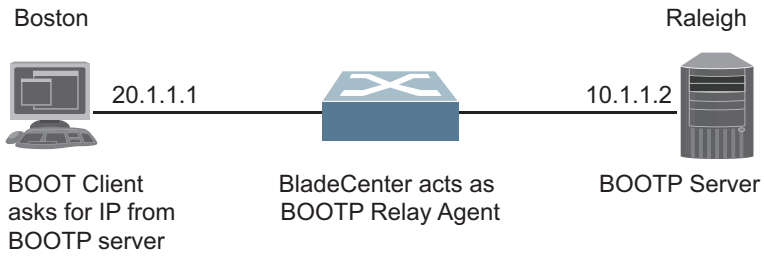
BOOTP Relay Agent

The CN4093 can function as a Bootstrap Protocol relay agent, enabling the switch to forward a client request for an IP address up to two BOOTP servers with IP addresses that have been configured on the switch.

When a switch receives a BOOTP request from a BOOTP client requesting an IP address, the switch acts as a proxy for the client. The request is then forwarded as a UDP Unicast MAC layer message to two BOOTP servers whose IP addresses are configured on the switch. The servers respond to the switch with a Unicast reply that contains the default gateway and IP address for the client. The switch then forwards this reply back to the client.

Figure 41 shows a basic BOOTP network example.

Figure 41. BOOTP Relay Agent Configuration



The use of two servers provide failover redundancy. The client request is forwarded to both BOOTP servers configured on the switch. However, no health checking is supported.

BOOTP Relay Agent Configuration

To enable the CN4093 to be the BOOTP forwarder, you need to configure the BOOTP server IP addresses on the switch, and enable BOOTP relay on the interface(s) on which the BOOTP requests are received.

Generally, you should configure the command on the switch IP interface that is closest to the client, so that the BOOTP server knows from which IP subnet the newly allocated IP address should come.

Use the following commands to configure the switch as a BOOTP relay agent:

```
CN 4093(config)# ip bootp-relay enable
CN 4093(config)# ip bootp-relay server <1-5> address <IPv4 address>
```

Use the following command to enable the Relay functionality on an IP interface:

```
CN 4093(config)# interface ip <interface number>
CN 4093(config-ip-if)# relay
CN 4093(config-ip-if)# exit
```

Domain-Specific BOOTP Relay Agent Configuration

Use the following commands to configure up to four domain-specific BOOTP relay agents for each of up to 10 VLANs:

```
CN 4093(config)# ip bootp-relay bcast-domain <1-10> vlan <VLAN number>
CN 4093(config)# ip bootp-relay bcast-domain <1-10> server <1-5> address
<IPv4 address>
CN 4093(config)# ip bootp-relay bcast-domain <1-10> enable
```

As with global relay agent servers, domain-specific BOOTP/DHCP functionality may be assigned on a per-interface basis.

Dynamic Host Configuration Protocol

Dynamic Host Configuration Protocol (DHCP) is a transport protocol that provides a framework for automatically assigning IP addresses and configuration information to other IP hosts or clients in a large TCP/IP network. Without DHCP, the IP address must be entered manually for each network device. DHCP allows a network administrator to distribute IP addresses from a central point and automatically send a new IP address when a device is connected to a different place in the network.

DHCP is an extension of another network IP management protocol, Bootstrap Protocol (BOOTP), with an additional capability of being able to dynamically allocate reusable network addresses and configuration parameters for client operation.

Built on the client/server model, DHCP allows hosts or clients on an IP network to obtain their configurations from a DHCP server, thereby reducing network administration. The most significant configuration the client receives from the server is its required IP address; (other optional parameters include the "generic" file name to be booted, the address of the default gateway, and so forth).

DHCP relay agent eliminates the need to have DHCP/BOOTP servers on every subnet. It allows the administrator to reduce the number of DHCP servers deployed on the network and to centralize them. Without the DHCP relay agent, there must be at least one DHCP server deployed at each subnet that has hosts needing to perform the DHCP request.

Note: The switch accepts gateway configuration parameters if they were not configured manually. The switch ignores DHCP gateway parameters if the gateway is configured.

DHCP Relay Agent

DHCP is described in RFC 2131, and the DHCP relay agent supported on CN4093s is described in RFC 1542. DHCP uses UDP as its transport protocol. The client sends messages to the server on port 67 and the server sends messages to the client on port 68.

DHCP defines the methods through which clients can be assigned an IP address for a finite lease period and allowing reassignment of the IP address to another client later. Additionally, DHCP provides the mechanism for a client to gather other IP configuration parameters it needs to operate in the TCP/IP network.

In the DHCP environment, the CN4093 acts as a relay agent. The DHCP relay feature enables the switch to forward a client request for an IP address to two BOOTP servers with IP addresses that have been configured on the switch.

When a switch receives a UDP broadcast on port 67 from a DHCP client requesting an IP address, the switch acts as a proxy for the client, replacing the client source IP (SIP) and destination IP (DIP) addresses. The request is then forwarded as a UDP Unicast MAC layer message to two BOOTP servers whose IP addresses are configured on the switch. The servers respond as a UDP Unicast message back to the switch, with the default gateway and IP address for the client. The destination

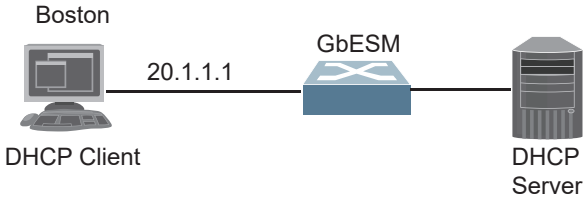
IP address in the server response represents the interface address on the switch that received the client request. This interface address tells the switch on which VLAN to send the server response to the client.

DHCP Relay Agent Configuration

To enable the CN4093 to be the BOOTP forwarder, you need to configure the DHCP/BOOTP server IP addresses on the switch. Generally, you should configure the switch IP interface on the client side to match the client's subnet, and configure VLANs to separate client and server subnets. The DHCP server knows from which IP subnet the newly allocated IP address should come.

The following figure shows a basic DHCP network example:

Figure 42. DHCP Relay Agent Configuration



In CN4093 implementation, there is no need for primary or secondary servers. The client request is forwarded to the BOOTP servers configured on the switch. The use of two servers provide failover redundancy. However, no health checking is supported.

Use the following commands to configure the switch as a DHCP relay agent:

```
CN 4093(config)# ip bootp-relay server 1 <IP address>
CN 4093(config)# ip bootp-relay server 2 <IP address>
CN 4093(config)# ip bootp-relay server 3 <IP address>
CN 4093(config)# ip bootp-relay server 4 <IP address>
CN 4093(config)# ip bootp-relay server 5 <IP address>
CN 4093(config)# ip bootp-relay enable
CN 4093(config)# show ip bootp-relay
```

Additionally, DHCP Relay functionality can be assigned on a per interface basis. Use the following command to enable the Relay functionality:

```
CN 4093(config)# interface ip <Interface number>
CN 4093(config-ip-if)# relay
```

Chapter 24. Internet Protocol Version 6

Internet Protocol version 6 (IPv6) is a network layer protocol intended to expand the network address space. IPv6 is a robust and expandable protocol that meets the need for increased physical address space. The switch supports the following RFCs for IPv6-related features:

- RFC 1981
- RFC 2404
- RFC 2410
- RFC 2451
- RFC 2460
- RFC 2461
- RFC 2462
- RFC 2474
- RFC 2526
- RFC 2711
- RFC 2740
- RFC 3289
- RFC 3306
- RFC 3307
- RFC 3411
- RFC 3412
- RFC 3413
- RFC 3414
- RFC 3484
- RFC 3602
- RFC 3810
- RFC 3879
- RFC 4007
- RFC 4213
- RFC 4291
- RFC 4293
- RFC 4293
- RFC 4301
- RFC 4302
- RFC 4303
- RFC 4306
- RFC 4307
- RFC 4443
- RFC 4552
- RFC 4718
- RFC 4835
- RFC 4861
- RFC 4862
- RFC 5095
- RFC 5114

This chapter describes the basic configuration of IPv6 addresses and how to manage the switch via IPv6 host management.

IPv6 Limitations

The following IPv6 features are not supported in this release.

- Dynamic Host Control Protocol for IPv6 (DHCPv6)
- Border Gateway Protocol for IPv6 (BGP)
- Routing Information Protocol for IPv6 (RIPng)

Most other Enterprise NOS 8.4 features permit IP addresses to be configured using either IPv4 or IPv6 address formats. However, the following switch features support IPv4 only:

- Default switch management IP address
- Bootstrap Protocol (BOOTP) and DHCP
- RADIUS, TACACS+ and LDAP
- QoS metering and re-marking ACLs for out-profile traffic
- VMware Virtual Center (vCenter) for VMready
- Routing Information Protocol (RIP)
- Internet Group Management Protocol (IGMP)
- Border Gateway Protocol (BGP)
- Virtual Router Redundancy Protocol (VRRP)
- sFLOW

IPv6 Address Format

The IPv6 address is 128 bits (16 bytes) long and is represented as a sequence of eight 16-bit hex values, separated by colons.

Each IPv6 address has two parts:

- Subnet prefix representing the network to which the interface is connected
- Local identifier, either derived from the MAC address or user-configured

The preferred hexadecimal format is as follows:

```
xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx:xxxx
```

Example IPv6 address:

```
FEDC:BA98:7654:BA98:FEDC:1234:ABCD:5412
```

Some addresses can contain long sequences of zeros. A single contiguous sequence of zeros can be compressed to :: (two colons). For example, consider the following IPv6 address:

```
FE80:0:0:0:2AA:FF:FA:4CA2
```

The address can be compressed as follows:

```
FE80::2AA:FF:FA:4CA2
```

Unlike IPv4, a subnet mask is not used for IPv6 addresses. IPv6 uses the subnet prefix as the network identifier. The prefix is the part of the address that indicates the bits that have fixed values or are the bits of the subnet prefix. An IPv6 prefix is written in address/prefix-length notation. For example, in the following address, 64 is the network prefix:

```
21DA:D300:0000:2F3C::/64
```

IPv6 addresses can be either user-configured or automatically configured. Automatically configured addresses always have a 64-bit subnet prefix and a 64-bit interface identifier. In most implementations, the interface identifier is derived from the switch's MAC address, using a method called EUI-64.

Most Enterprise NOS 8.4 features permit IP addresses to be configured using either IPv4 or IPv6 address formats. Throughout this manual, *IP address* is used in places where either an IPv4 or IPv6 address is allowed. In places where only one type of address is allowed, the type (*IPv4* or *IPv6* is specified).

IPv6 Address Types

IPv6 supports three types of addresses: unicast (one-to-one), multicast (one-to-many), and anycast (one-to-nearest). Multicast addresses replace the use of broadcast addresses.

Unicast Address

Unicast is a communication between a single host and a single receiver. Packets sent to a unicast address are delivered to the interface identified by that address. IPv6 defines the following types of unicast address:

- Global Unicast address: An address that can be reached and identified globally. Global Unicast addresses use the high-order bit range up to FF00, therefore all non-multicast and non-link-local addresses are considered to be global unicast. A manually configured IPv6 address must be fully specified. Auto-configured IPv6 addresses are comprised of a prefix combined with the 64-bit EUI. RFC 4291 defines the IPv6 addressing architecture.

  The interface ID must be unique within the same subnet.

- Link-local unicast address: An address used to communicate with a neighbor on the same link. Link-local addresses use the format `FE80::EUI`

  Link-local addresses are designed to be used for addressing on a single link for purposes such as automatic address configuration, neighbor discovery, or when no routers are present.

  Routers must not forward any packets with link-local source or destination addresses to other links.

Multicast Address

Multicast is communication between a single host and multiple receivers. Packets are sent to all interfaces identified by that address. An interface may belong to any number of multicast groups.

A multicast address (FF00 - FFFF) is an identifier for a group interface. The multicast address most often encountered is a solicited-node multicast address using prefix `FF02::1:FF00:0000/104` with the low-order 24 bits of the unicast or anycast address.

The following well-known multicast addresses are pre-defined. The group IDs defined in this section are defined for explicit scope values, as follows:

`FF00:::::::0` through `FF0F:::::::0`

Anycast Address

Packets sent to an anycast address or list of addresses are delivered to the nearest interface identified by that address. Anycast is a communication between a single sender and a list of addresses.

Anycast addresses are allocated from the unicast address space, using any of the defined unicast address formats. Thus, anycast addresses are syntactically indistinguishable from unicast addresses. When a unicast address is assigned to more than one interface, thus turning it into an anycast address, the nodes to which the address is assigned must be explicitly configured to know that it is an anycast address.

IPv6 Address Auto-configuration

IPv6 supports the following types of address auto-configuration:

- **Stateful address configuration**

  Address configuration is based on the use of a stateful address configuration protocol, such as DHCPv6, to obtain addresses and other configuration options.

- **Stateless address configuration**

  Address configuration is based on the receipt of Router Advertisement messages that contain one or more Prefix Information options.

Enterprise NOS 8.4 supports stateless address configuration.

Stateless address configuration allows hosts on a link to configure themselves with link-local addresses and with addresses derived from prefixes advertised by local routers. Even if no router is present, hosts on the same link can configure themselves with link-local addresses and communicate without manual configuration.

IPv6 Interfaces

Each IPv6 interface supports multiple IPv6 addresses. You can manually configure up to two IPv6 addresses for each interface, or you can allow the switch to use stateless autoconfiguration. By default, the switch automatically configures the IPv6 address of its management interface.

You can manually configure two IPv6 addresses for each interface, as follows:

- Initial IPv6 address is a global unicast or anycast address .

```
CN 4093(config)# interface ip <interface number>
CN 4093(config-ip-if)# ipv6 address <IPv6 address>
```

  Note that you cannot configure both addresses as anycast. If you configure an anycast address on the interface you must also configure a global unicast address on that interface.

- Second IPv6 address can be a unicast or anycast address .

```
CN 4093(config-ip-if)# ipv6 secaddr6 <IPv6 address>
CN 4093(config-ip-if)# exit
```

You cannot configure an IPv4 address on an IPv6 management interface. Each interface can be configured with only one address type: either IPv4 or IPv6, but not both. When changing between IPv4 and IPv6 address formats, the prior address settings for the interface are discarded.

Each IPv6 interface can belong to only one VLAN. Each VLAN can support only one IPv6 interface. Each VLAN can support multiple IPv4 interfaces.

Interface 125/126 is reserved for IPv6 host support. This interface is included in management VLAN 4095. Use the following commands to configure the IPv6 gateway:

```
CN 4093(config)# ip gateway6 1 address <IPv6 address>
CN 4093(config)# ip gateway6 1 enable
```

IPv6 gateway 1 is reserved for IPv6 data interfaces. IPv6 gateway 3 and 4 are the default IPv6 management gateways.

Neighbor Discovery

The switch uses Neighbor Discovery protocol (ND) to gather information about other router and host nodes, including the IPv6 addresses. Host nodes use ND to configure their interfaces and perform health detection. ND allows each node to determine the link-layer addresses of neighboring nodes, and to keep track of each neighbor's information. A neighboring node is a host or a router that is linked directly to the switch. The switch supports Neighbor Discovery as described in RFC 4861.

Neighbor Discover messages allow network nodes to exchange information, as follows:

- *Neighbor Solicitations* allow a node to discover information about other nodes.
- *Neighbor Advertisements* are sent in response to Neighbor Solicitations. The Neighbor Advertisement contains information required by nodes to determine the link-layer address of the sender, and the sender's role on the network.
- IPv6 hosts use *Router Solicitations* to discover IPv6 routers. When a router receives a Router Solicitation, it responds immediately to the host.
- Routers uses *Router Advertisements* to announce its presence on the network, and to provide its address prefix to neighbor devices. IPv6 hosts listen for Router Advertisements, and uses the information to build a list of default routers. Each host uses this information to perform autoconfiguration of IPv6 addresses.
- *Redirect messages* are sent by IPv6 routers to inform hosts of a better first-hop address for a specific destination. Redirect messages are only sent by routers for unicast traffic, are only unicast to originating hosts, and are only processed by hosts.

ND configuration for various advertisements, flags, and interval settings is performed on a per-interface basis using the following command path:

```
CN 4093(config)# interface ip <interface number>
CN 4093(config-ip-if)# [no] ipv6 nd ?
CN 4093(config-ip-if)# exit
```

To add or remove entries in the static neighbor cache, use the following command path:

```
CN 4093(config)# [no] ip neighbors ?
```

Host vs. Router

Each IPv6 interface can be configured as a router node or a host node, as follows:

- A router node's IP address is configured manually. Router nodes can send Router Advertisements.
- A host node's IP address is auto-configured. Host nodes listen for Router Advertisements that convey information about devices on the network.

Note: When IP forwarding is turned on, all IPv6 interfaces configured on the switch can forward packets.

You can configure each IPv6 interface as either a host node or a router node. You can manually assign an IPv6 address to an interface in host mode, or the interface can be assigned an IPv6 address by an upstream router, using information from router advertisements to perform stateless auto-configuration.

To set an interface to host mode, use the following command:

```
CN 4093(config)# interface ip <interface number>
CN 4093(config-ip-if)# ip6host
CN 4093(config-ip-if)# exit
```

By default, host mode is enabled on the management interface, and disabled on data interfaces.

The CN4093 supports up to 1156 IPv6 routes.

Supported Applications

The following applications have been enhanced to provide IPv6 support.

- **Ping**

  The **ping** command supports IPv6 addresses. Use the following format to ping an IPv6 address:

  ping *<host name>*|*<IPv6 address>* [**-n** *<tries (0-4294967295)>*] [**-w** *<msec delay (0-4294967295)>*] [**-l** *<length (0/32-65500/2080)>*] [**-s** *<IP source>*] [**-v** *<TOS (0-255)>*] [**-f**] [**-t**]

  To ping a link-local address (begins with FE80), provide an interface index, as follows:

  ping *<IPv6 address>***%***<Interface index>* [**-n** *<tries (0-4294967295)>*] [**-w** *<msec delay (0-4294967295)>*] [**-l** *<length (0/32-65500/2080)>*] [**-s** *<IP source>*] [**-v** *<TOS (0-255)>*] [**-f**] [**-t**]

- **Traceroute**

  The **traceroute** command supports IPv6 addresses (but not link-local addresses).
  Use the following format to perform a traceroute to an IPv6 address:

  traceroute *<host name>*| *<IPv6 address>* [*<max-hops (1-32)>* [*<msec delay (1-4294967295)>*]]

- **Telnet server**

  The **telnet** command supports IPv6 addresses, but not link-local addresses. Use the following format to Telnet into an IPv6 interface on the switch:

  telnet *<host name>*| *<IPv6 address>* [*<port>*]

- **Telnet client**

  The **telnet** command supports IPv6 addresses, but not link-local addresses. Use the following format to Telnet to an IPv6 address:

  telnet *<host name>*| *<IPv6 address>* [*<port>*]

- **HTTP/HTTPS**

  The HTTP/HTTPS servers support both IPv4 and IPv6 connections.

- **SSH**

  Secure Shell (SSH) connections over IPv6 are supported, but not link-local addresses. The following syntax is required from the client:

  ssh -u *<IPv6 address>*

  Example:

  ssh -u 2001:2:3:4:0:0:0:142

- **TFTP**

  The TFTP commands support both IPv4 and IPv6 addresses. Link-local addresses are not supported.

- **FTP**

  The FTP commands support both IPv4 and IPv6 addresses. Link-local addresses are not supported.

- **DNS client**

  DNS commands support both IPv4 and IPv6 addresses. Link-local addresses are not supported. Use the following command to specify the type of DNS query to be sent first:

```
CN 4093(config)# ip dns ipv6 request-version {ipv4|ipv6}
```

  If you set the request version to `ipv4`, the DNS application sends an *A* query first, to resolve the hostname with an IPv4 address. If no *A* record is found for that hostname (no IPv4 address for that hostname) an *AAAA* query is sent to resolve the hostname with a IPv6 address.

  If you set the request version to `ipv6`, the DNS application sends an *AAAA* query first, to resolve the hostname with an IPv6 address. If no *AAAA* record is found for that hostname (no IPv6 address for that hostname) an *A* query is sent to resolve the hostname with an IPv4 address.

IPv6 Configuration

This section provides steps to configure IPv6 on the switch.

Configuration Guidelines

When you configure an interface for IPv6, consider the following guidelines:

- Support for subnet router anycast addresses is not available.
- Interface 125/126 are reserved for IPv6 management.
- A single interface can accept either IPv4 or IPv6 addresses, but not both IPv4 and IPv6 addresses.
- A single interface can accept multiple IPv6 addresses.
- A single interface can accept only one IPv4 address.
- If you change the IPv6 address of a configured interface to an IPv4 address, all IPv6 settings are deleted.
- A single VLAN can support only one IPv6 interface.
- Health checks are not supported for IPv6 gateways.
- IPv6 interfaces support Path MTU Discovery. The CPU's MTU is fixed at 1500 bytes.
- Support for jumbo frames (1,500 to 9,216 byte MTUs) is limited. Any jumbo frames intended for the CPU must be fragmented by the remote node. The switch can re-assemble fragmented packets up to 9k. It can also fragment and transmit jumbo packets received from higher layers.

IPv6 Configuration Examples

IPv6 Configuration Example 1

The following example uses IPv6 host mode to autoconfigure an IPv6 address for the interface. By default, the interface is assigned to VLAN 1.

1. Enable IPv6 host mode on an interface.

```
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip6host
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
```

2. Configure the IPv6 default gateway.

```
CN 4093(config)# ip gateway6 1 address 2001:BA98:7654:BA98:FEDC:1234:
ABCD:5412
CN 4093(config)# ip gateway6 1 enable
```

3. Verify the interface address.

```
CN 4093(config)# show interface ip 2
```

IPv6 Configuration Example 2

Use the following example to manually configure IPv6 on an interface.

1. Assign an IPv6 address and prefix length to the interface.

```
CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ipv6 address 2001:BA98:7654:BA98:FEDC:1234:
ABCD:5214
CN 4093(config-ip-if)# ipv6 prefixlen 64
CN 4093(config-ip-if)# ipv6 seccaddr6 2003::1 32
CN 4093(config-ip-if)# vlan 2
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
```

   The secondary IPv6 address is compressed, and the prefix length is 32.

2. Configure the IPv6 default gateway.

```
CN 4093(config)# ip gateway6 1 address 2001:BA98:7654:BA98:FEDC:1234:
ABCD:5412
CN 4093(config)# ip gateway6 1 enable
```

3. Configure Router advertisements for the interface (optional)

```
CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# no ipv6 nd suppress-ra
```

4. Verify the configuration.

```
CN 4093(config-ip-if)# show layer3
```

Chapter 25. Using IPsec with IPv6

Internet Protocol Security (IPsec) is a protocol suite for securing Internet Protocol (IP) communications by authenticating and encrypting each IP packet of a communication session. IPsec also includes protocols for establishing mutual authentication between agents at the beginning of the session and negotiation of cryptographic keys to be used during the session.

Since IPsec was implemented in conjunction with IPv6, all implementations of IPv6 must contain IPsec. To support the National Institute of Standards and Technology (NIST) recommendations for IPv6 implementations, Enterprise NOS IPv6 feature compliance has been extended to include the following IETF RFCs, with an emphasis on IP Security (IPsec) and Internet Key Exchange version 2, and authentication/confidentiality for OSPFv3:

- RFC 4301 for IPv6 security
- RFC 4302 for the IPv6 Authentication Header
- RFCs 2404, 2410, 2451, 3602, and 4303 for IPv6 Encapsulating Security Payload (ESP), including NULL encryption, CBC-mode 3DES and AES ciphers, and HMAC-SHA-1-96.
- RFCs 4306, 4307, 4718, and 4835 for IKEv2 and cryptography
- RFC 4552 for OSPFv3 IPv6 authentication
- RFC 5114 for Diffie-Hellman groups

Note: This implementation of IPsec supports DH groups 1, 2, 5, 14, and 24.

The following topics are discussed in this chapter:

IPsec Protocols

The Enterprise NOS implementation of IPsec supports the following protocols:

- Authentication Header (AH)

  AHs provide connectionless integrity and data origin authentication for IP packets, and provide protection against replay attacks. In IPv6, the AH protects the AH itself, the Destination Options extension header after the AH, and the IP payload. It also protects the fixed IPv6 header and all extension headers before the AH, except for the mutable fields DSCP, ECN, Flow Label, and Hop Limit. AH is defined in RFC 4302.

- Encapsulating Security Payload (ESP)

  ESPs provide confidentiality, data origin authentication, integrity, an anti-replay service (a form of partial sequence integrity), and some traffic flow confidentiality. ESPs may be applied alone or in combination with an AH. ESP is defined in RFC 4303.

- Internet Key Exchange Version 2 (IKEv2)

  IKEv2 is used for mutual authentication between two network elements. An IKE establishes a security association (SA) that includes shared secret information to efficiently establish SAs for ESPs and AHs, and a set of cryptographic algorithms to be used by the SAs to protect the associated traffic. IKEv2 is defined in RFC 4306.

Using IKEv2 as the foundation, IPsec supports ESP for encryption and/or authentication, and/or AH for authentication of the remote partner.

Both ESP and AH rely on security associations. A security association (SA) is the bundle of algorithms and parameters (such as keys) that encrypt and authenticate a particular flow in one direction.

Using IPsec with the CN4093

IPsec supports the fragmentation and reassembly of IP packets that occurs when data goes to and comes from an external device. The Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch acts as an end node that processes any fragmentation and reassembly of packets but does not forward the IPsec traffic. The IKEv2 key must be authenticated before you can use IPsec.

The security protocol for the session key is either ESP or AH. Outgoing packets are labeled with the SA SPI (Security Parameter Index), which the remote device will use in its verification and decryption process.

Every outgoing IPv6 packet is checked against the IPsec policies in force. For each outbound packet, after the packet is encrypted, the software compares the packet size with the MTU size that it either obtains from the default minimum maximum transmission unit (MTU) size (1500) or from path MTU discovery. If the packet size is larger than the MTU size, the receiver drops the packet and sends a message containing the MTU size to the sender. The sender then fragments the packet into smaller pieces and retransmits them using the correct MTU size.

The maximum traffic load for each IPSec packet is limited to the following:

- IKEv2 SAs: 5
- IPsec SAs: 10 (5 SAs in each direction)
- SPDs: 20 (10 policies in each direction)

IPsec is implemented as a software cryptography engine designed for handling control traffic, such as network management. IPsec is not designed for handling data traffic, such as a VPN.

Setting up Authentication

Before you can use IPsec, you need to have key policy authentication in place. There are two types of key policy authentication:

- Preshared key (default)

  The parties agree on a shared, secret key that is used for authentication in an IPsec policy. During security negotiation, information is encrypted before transmission by using a session key created by using a Diffie-Hellman calculation and the shared, secret key. Information is decrypted on the receiving end using the same key. One IPsec peer authenticates the other peer's packet by decryption and verification of the hash inside the packet (the hash inside the packet is a hash of the preshared key). If authentication fails, the packet is discarded.

- Digital certificate (using RSA algorithms)

  The peer being validated must hold a digital certificate signed by a trusted Certificate Authority and the private key for that digital certificate. The side performing the authentication only needs a copy of the trusted certificate authorities digital certificate. During IKEv2 authentication, the side being validated sends a copy of the digital certificate and a hash value signed using the private key. The certificate can be either generated or imported.

Note: During the IKEv2 negotiation phase, the digital certificate takes precedence over the preshared key.

Creating an IKEv2 Proposal

With IKEv2, a single policy can have multiple encryption and authentication types, as well as multiple integrity algorithms.

To create an IKEv2 proposal:

1. Enter IKEv2 proposal mode.

```
CN 4093(config)# ikev2 proposal
```

2. Set the DES encryption algorithm.

```
CN 4093(config-ikev2-prop)# encryption {3des|aes-cbc|des}    (default: 3des)
```

3. Set the authentication integrity algorithm type.

```
CN 4093(config-ikev2-prop)# integrity {md5|sha1}             (default: sha1)
```

4. Set the Diffie-Hellman group.

```
CN 4093(config-ikev2-prop)# group {1|2|5|14|24}              (default: 2)
```

Importing an IKEv2 Digital Certificate

To import an IKEv2 digital certificate for authentication:

1. Import the CA certificate file.

```
CN 4093(config)# copy tftp ca-cert address <hostname or IPv4 address>

Source file name: <path and filename of CA certificate file>
Port type ["DATA"/"MGT"]: >
Confirm download operation [y/n]: y
```

2. Import the host key file.

```
CN 4093(config)# copy tftp host-key address <hostname or IPv4 address>

Source file name: <path and filename of host private key file>
Port type ["DATA"/"MGT"]: >
Confirm download operation [y/n]: y
```

3. Import the host certificate file.

```
CN 4093(config)# copy tftp host-cert address <hostname or IPv4 address>

Source file name: <path and filename of host certificate file>
Port type ["DATA"/"MGT"]: >
Confirm download operation [y/n]: y
```

Note: When prompted for the port to use for download the file, if you used a management port to connect the switch to the server, enter **`mgt`**, otherwise enter **`data`**.

Generating a Certificate Signing Request

Before a digital certificate can be signed by a Certificate Authority (CA), it needs to be created. The generation of a certificate involves creating a Certificate Signing Request (CSR). The CSR includes various information related to the device and a public key. The public key is included in the CSR file itself and the private key associated with the public key is generated separately and kept private. The CSR can then be exported to a remote device to be signed by a CA.

1. Create an HTTPS CSR defining the information you want to be used in the various fields:

```
CN 4093(config)# access https generate-csr
Country Name (2 letter code) []:  <country code>
State or Province Name (full name) []:  <state>
Locality Name (eg, city) []:  <city>
Organization Name (eg, company) []:  <company>
Organizational Unit Name (eg, section) []:  <org. unit>
Common Name (eg, YOUR name) []:  <name>
Email (eg, email address) []:  <email address>
Confirm Generate CSR? [y/n]: y
.........................................+++
.............+++
Cert Req generated successfully
```

2. To verify the CSR you can use the following command:

show https host-csr [pem-format|txt-format]

```
CN 4093> show https host-csr txt-format

Certificate Request:
    Data:
        Version: 0 (0x0)
        Subject: C=US, ST=Cali, L=Santa Barbara, O=Lenovo, OU=Sales, CN=www.zagat.com
        Subject Public Key Info:
            Public Key Algorithm: rsaEncryption
            RSA Public Key: (2048 bit)
                Modulus (2048 bit):
                    00:b5:05:f6:d5:ad:ab:f2:1d:a9:57:c4:bc:84:1b:
                    c6:bc:cd:04:95:ea:ad:ec:4a:44:3a:6e:42:9f:39:
                    96:14:11:a7:8e:3e:6f:da:9a:42:c6:c4:62:a1:33:
                    0e:a8:d3:6a:21:ce:f3:3c:4f:c1:8d:d1:e7:9e:c7:
                    29:04:ea:c6:7d:54:9a:4e:10:24:10:38:45:c6:4b:
                    13:19:f2:dd:8a:83:3f:5c:cf:8b:85:a7:2a:b0:eb:
                    7a:26:1f:4c:94:47:01:81:6a:59:d5:f5:d6:7e:3b:
                    b5:bc:e4:3f:6d:dd:84:15:07:61:93:e0:d1:40:f8:
                    9d:15:d0:a6:e1:9b:a4:ab:85:b5:2b:f0:56:e9:ef:
                    36:43:2b:aa:be:1b:63:3c:fd:74:ab:78:76:53:12:
                    e6:65:4c:0d:07:91:df:b3:91:96:f4:55:f7:37:73:
                    8c:f6:77:d7:9d:2b:a5:bd:17:3f:11:f2:85:4b:d6:
                    b4:1d:3f:70:1f:13:bb:5e:2e:4c:a8:ad:6a:7f:11:
                    36:97:a6:25:0a:87:66:31:c9:92:59:03:31:5d:ff:
                    df:c6:aa:93:7c:51:9f:8e:1b:6f:2a:be:c4:4c:66:
                    d6:2c:4b:6d:e6:ae:4e:02:82:fc:fa:a1:de:3b:c9:
                    24:25:d5:6e:15:15:18:ce:9b:a6:98:ad:0c:32:1f:
                    94:01
                Exponent: 65537 (0x10001)
        Attributes:
            a0:00
    Signature Algorithm: sha256WithRSAEncryption
        24:26:dd:96:49:47:9d:78:74:48:9b:63:4c:32:f0:78:da:7d:
        82:c9:17:6d:7e:93:38:60:94:d5:02:c1:31:dc:42:69:f5:57:
        46:a8:44:5a:99:ea:55:d3:99:bf:f0:48:3b:ef:60:fd:50:e6:
        33:cd:89:86:d3:51:97:f2:d1:68:6f:88:8c:e7:0f:3e:19:2a:
        f4:ea:6b:dc:05:24:d7:98:cd:a3:d3:c3:ef:03:93:8b:3f:fe:
        75:5e:67:f1:48:b6:20:a6:ff:ae:5a:25:41:7f:e4:c8:48:d4:
        63:37:16:98:9e:2d:1b:b6:65:7a:0d:90:87:07:19:f0:02:17:
        3a:3e:fd:f0:40:3e:a4:0f:53:97:9b:d5:18:22:78:f3:07:94:
        63:be:f9:f2:5c:23:6d:0f:22:d1:17:db:38:24:5c:6b:7b:e0:
        41:a6:51:28:30:2c:f4:1d:62:6c:06:f2:4c:0c:5b:79:51:13:
        73:f8:88:ba:2e:05:98:5d:41:5e:9d:58:b1:0c:8f:fc:f2:79:
        d5:30:7c:95:e9:ff:9a:cc:dd:d9:4c:2e:98:32:5a:ab:cd:59:
        a4:37:a5:38:03:4e:e7:27:dc:14:c8:75:9d:ca:e0:62:37:02:
        19:17:16:e3:92:c0:c3:16:13:26:c9:40:d7:ec:f2:8c:8e:fc:
        1a:dc:27:4c
```

```
CN 4093> show https host-csr pem-format

-----BEGIN CERTIFICATE REQUEST-----
MIICtDCCAZwCAQAwbzELMAkGA1UEBhMCVVMxEzARBgNVBAgMCkNhbGlmb3JuaWEx
ETAPBgNVBAcMCFNhbiBKb3NlMQwwCgYDVQQKDANBQkMxFDASBgNVBAsMC0VuZ2lu
ZWVyaW5nMRQwEgYDVQQDDAt3d3cuYWJjLmNvbTCCASIwDQYJKoZIhvcNAQEBBQAD
ggEPADCCAQoCggEBAMEnVJBSnIY9OGiCcH+xmYKpWga7E5j9JSK9JU57Md7NofJ2
FvQ8hfPO8b4bzLQzKbNBxGc59BJjZJ5w8eGKRDCjlIf1uIAgg3Gs8ZK1FozOUJZN
xbtYBx6QrTBYmXdHStQ7CQ9sfWhnEnusnvc8bxNlukyuEcFsAUdz93r1sEfN3cDe
/bO43l7GmvhTEdmfFvAfgi9b9RDqUjliATpgS3+a2kwhjvHCTeveQN1/MYQZvbJo
V4qq+pgQOt9ZJOMDrGQ0YmO1p84+GdxXVwGePCOvCRLESsq5rQb3zPSVvWnTsq0G
gURvbV+VQN9dI9lANZGZJi6BRNIRdBen/dH0KRcCAwEAAaAAMA0GCSqGSIb3DQEB
BQUAA4IBAQCSLDOrOnl7kaZri2OjDpzgiiG+9Skde3MehaklddfZnCkT1ALL3ZXY
xWwYnvF5jAgnHhxRJbPOzwHNDWMtZiiNOTHyzHVptsyRBv70Kb8odJmuyKWDqunJ
Ho1hHe63a7MRLFkQ+6io3kGrmq1bdM5U6xvvS+0ZXXUaiK1p/lNLOrsYk45D01Az
YHhcdRQtFUbQxqbirpi0jLsi82X7JCNQ2XCP6dhphkWKI6wsCvmlJdazW+V/gH/X
wqMkNF8mkodzlhC+1COd2yzSXxqpG/Xf0TRF9SAyN5vK4NDDzZu+iPvh6RkXXeNV
neyr2J5JENyGORPynuVoHWuzEy+5GUHa
-----END CERTIFICATE REQUEST-----
```

3. Export the CSR file to an external server:

```
CN 4093(config)# copy cert-request tftp

Port type ["DATA"/"MGT"/"EXTM"]: <port type>
Address or name of remote host: <hostname or IPv4 address>
Destination file name: <path and filename on the remote server>

Certificate request successfully tftp'd to...
```

Generating an IKEv2 Digital Certificate

To create an IKEv2 digital certificate for authentication:

1. Create an HTTPS certificate defining the information you want to be used in the various fields.

```
CN 4093(config)# access https generate-certificate
Country Name (2 letter code) []:  <country code>
State or Province Name (full name) []:  <state>
Locality Name (eg, city) []:  <city>
Organization Name (eg, company) []:  <company>
Organizational Unit Name (eg, section) []:  <org. unit>
Common Name (eg, YOUR name) []:  <name>
Email (eg, email address) []:  <email address>
Confirm generat‘eywing certificate? [y/n]: y
Generating certificate. Please wait (approx 30 seconds)
restarting SSL agent
```

2. Save the HTTPS certificate.

   The certificate is valid only until the switch is rebooted. To save the certificate so that it is retained beyond reboot or power cycles, use the following command:

```
CN 4093(config)# access https save-certificate
```

3. Enable IKEv2 RSA-signature authentication:

```
CN 4093(config)# access https enable
```

Enabling IKEv2 Preshared Key Authentication

To set up IKEv2 preshared key authentication:

1. Enter the local preshared key.

```
CN 4093(config)# ikev2 preshare-key local <preshared key, a string of 1-256 chars>
```

2. If asymmetric authentication is supported, enter the remote key:

```
CN 4093(config)# ikev2 preshare-key remote <preshared key> <IPv6 host>
```

where the following parameters are used:

- *preshared key* — A string of 1-256 characters
- *IPv6 host* — An IPv6-format host, such as "3000::1"

3. Set up the IKEv2 identification type by entering *one* of the following commands:

```
CN 4093(config)# ikev2 identity local address (use an IPv6 address)
CN 4093(config)# ikev2 identity local email <email address>
CN 4093(config)# ikev2 identity local fqdn <domain name>
```

To disable IKEv2 RSA-signature authentication method and enable preshared key authentication, enter:

```
CN 4093(config)# no access https
```

Setting Up a Key Policy

When configuring IPsec, you must define a key policy. This key policy can be either manual or dynamic. Either way, configuring a policy involves the following steps:

- Create a transform set—This defines which encryption and authentication algorithms are used.
- Create a traffic selector—This describes the packets to which the policy applies.
- Establish an IPsec policy.
- Apply the policy.

1. To define which encryption and authentication algorithms are used, create a transform set:

```
CN 4093(config)# ipsec transform-set <transform ID> <encryption method> <integrity algorithm> <AH authentication algorithm>
```

where the following parameters are used:

- *transform ID* — A number from 1-10
- *encryption method* — One of the following: **esp-des** | **esp-3des** | **esp-aes-cbc** | **esp-null**
- *integrity algorithm* — One of the following: **esp-sha1** | **esp-md5** | **none**
- *AH authentication algorithm* — One of the following: **ah-sha1** | **ah-md5** | **none**

2. Decide whether to use tunnel or transport mode. The default mode is transport.

```
CN 4093(config)# ipsec transform-set tunnel|transport
```

3. To describe the packets to which this policy applies, create a traffic selector using the following commands:

```
CN 4093(config)# ipsec traffic-selector <traffic selector number> {permit|deny}
{any|icmp {<ICMPv6 type>|any}|tcp} {<source IP address>|any} {<destination IP address>|
|any} [<prefix length>]
```

   where the following parameters are used:

   - *traffic selector number* — an integer from 1-10
   - **permit|deny** — whether or not to permit IPsec encryption of traffic that meets the criteria specified in this command
   - **proto/any** — apply the selector to any type of traffic
   - **proto/icmp** *type*|**any** — only apply the selector only to ICMP traffic of the specified *type* (an integer from 1-255) or to any ICMP traffic
   - **proto/tcp** — only apply the selector to TCP traffic
   - *source IP address*|**any** — the source IP address in IPv6 format or "any" source
   - *destination IP address*|**any** — the destination IP address in IPv6 format or "any" destination
   - *prefix length* — (Optional) the length of the destination IPv6 prefix; an integer from 1-128

   Permitted traffic that matches the policy in force is encrypted, while denied traffic that matches the policy in force is dropped. Traffic that does not match the policy bypasses IPsec and passes through *clear* (unencrypted).

4. Choose whether to use a manual or a dynamic policy.

Using a Manual Key Policy

A manual policy involves configuring policy and manual SA entries for local and remote peers.

To configure a manual key policy, you need:

- The IP address of the peer in IPv6 format (for example, "3000::1").
- Inbound/Outbound session keys for the security protocols.

You can then assign the policy to an interface. The peer represents the other end of the security association. The security protocol for the session key can be either ESP or AH.

To create and configure a manual policy:

1. Enter a manual policy to configure.

```
CN 4093(config)# ipsec manual-policy <policy number>
```

2. Configure the policy.

```
CN 4093(config-ipsec-manual)#peer <peer's IPv6 address>
CN 4093(config-ipsec-manual)#traffic-selector <IPsec traffic selector>
CN 4093(config-ipsec-manual)#transform-set <IPsec transform set>
CN 4093(config-ipsec-manual)#in-ah auth-key <inbound AH IPsec key>
CN 4093(config-ipsec-manual)#in-ah auth-spi <inbound AH IPsec SPI>
CN 4093(config-ipsec-manual)#in-esp cipher-key <inbound ESP cipher key>
CN 4093(config-ipsec-manual)#in-esp auth-spi <inbound ESP SPI>
CN 4093(config-ipsec-manual)#in-esp auth-key <inbound ESP authenticator key>
CN 4093(config-ipsec-manual)#out-ah auth-key <outbound AH IPsec key>
CN 4093(config-ipsec-manual)#out-ah auth-spi <outbound AH IPsec SPI>
CN 4093(config-ipsec-manual)#out-esp cipher-key <outbound ESP cipher key>
CN 4093(config-ipsec-manual)#out-esp auth-spi <outbound ESP SPI>
CN 4093(config-ipsec-manual)#out-esp auth-key <outbound ESP authenticator key>
```

where the following parameters are used:

- *peer's IPv6 address* — The IPv6 address of the peer (for example, 3000::1)
- *IPsec traffic-selector* — A number from1-10
- *IPsec of transform-set* — A number from1-10
- *inbound AH IPsec key* — The inbound AH key code, in hexadecimal
- *inbound AH IPsec SPI* — A number from 256-4294967295
- *inbound ESP cipher key* — The inbound ESP key code, in hexadecimal
- *inbound ESP SPI* — A number from 256-4294967295
- *inbound ESP authenticator key* — The inbound ESP authenticator key code, in hexadecimal
- *outbound AH IPsec key* — The outbound AH key code, in hexadecimal
- *outbound AH IPsec SPI* — A number from 256-4294967295
- *outbound ESP cipher key* — The outbound ESP key code, in hexadecimal
- *outbound ESP SPI* — A number from 256-4294967295

- *outbound ESP authenticator key* The outbound ESP authenticator key code, in hexadecimal

Note: When configuring a manual policy ESP, the ESP authenticator key is optional.

3. After you configure the IPSec policy, you need to apply it to the interface to enforce the security policies on that interface and save it to keep it in place after a reboot. To accomplish this, enter:

```
CN 4093(config-ip)# interface ip <IP interface number, 1-128>
CN 4093(config-ip-if)# address <IPv6 address>
CN 4093(config-ip-if)# ipsec manual-policy <policy index, 1-10>
CN 4093(config-ip-if)# enable (enable the IP interface)
CN 4093# write (save the current configuration)
```

Using a Dynamic Key Policy

When you use a dynamic key policy, the first packet triggers IKE and sets the IPsec SA and IKEv2 SA. The initial packet negotiation also determines the lifetime of the algorithm, or how long it stays in effect. When the key expires, a new key is automatically created. This helps prevent break-ins.

To configure a dynamic key policy:

1. Choose a dynamic policy to configure.

```
CN 4093(config)# ipsec dynamic-policy <policy number>
```

2. Configure the policy.

```
CN 4093(config-ipsec-dynamic)# peer <peer's IPv6 address>
CN 4093(config-ipsec-dynamic)# traffic-selector <index of traffic selector>
CN 4093(config-ipsec-dynamic)# transform-set <index of transform set>
CN 4093(config-ipsec-dynamic)# sa-lifetime <SA lifetime, in seconds>
CN 4093(config-ipsec-dynamic)# pfs enable|disable
```

   where the following parameters are used:

- *peer's IPv6 address* — The IPv6 address of the peer (for example, 3000::1)
- *index of traffic-selector* — A number from1-10
- *index of transform-set* — A number from1-10
- *SA lifetime, in seconds* — The length of time the SA is to remain in effect; an integer from120-86400
- **pfs enable|disable** — Whether to enable or disable the perfect forward security feature. The default is **disable**.

Note: In a dynamic policy, the AH and ESP keys are created by IKEv2.

3. After you configure the IPSec policy, you need to apply it to the interface to enforce the security policies on that interface and save it to keep it in place after a reboot. To accomplish this, enter:

```
CN 4093(config-ip)# interface ip <IP interface number, 1-128>
CN 4093(config-ip-if)# address <IPv6 address>
CN 4093(config-ip-if)# ipsec dynamic-policy <policy index, 1-10>
CN 4093(config-ip-if)# enable (enable the IP interface)
CN 4093# write (save the current configuration)
```

Chapter 26. Routing Information Protocol

In a routed environment, routers communicate with one another to keep track of available routes. Routers can learn about available routes dynamically using the Routing Information Protocol (RIP). Enterprise NOS software supports RIP version 1 (RIPv1) and RIP version 2 (RIPv2) for exchanging TCP/IPv4 route information with other routers.

Note: Enterprise NOS 8.4 does not support IPv6 for RIP.

Distance Vector Protocol

RIP is known as a distance vector protocol. The vector is the network number and next hop, and the distance is the cost associated with the network number. RIP identifies network reachability based on metric, and metric is defined as hop count. One hop is considered to be the distance from one switch to the next, which typically is 1.

When a switch receives a routing update that contains a new or changed destination network entry, the switch adds 1 to the metric value indicated in the update and enters the network in the routing table. The IPv4 address of the sender is used as the next hop.

Stability

RIP includes a number of other stability features that are common to many routing protocols. For example, RIP implements the split horizon and hold-down mechanisms to prevent incorrect routing information from being propagated.

RIP prevents routing loops from continuing indefinitely by limiting the number of hops allowed in a path from the source to a destination. The maximum number of hops in a path is 15. The network destination network is considered unreachable if increasing the metric value by 1 causes the metric to be 16 (that is infinity). This limits the maximum diameter of a RIP network to less than 16 hops.

RIP is often used in stub networks and in small autonomous systems that do not have many redundant paths.

Routing Updates

RIP sends routing-update messages at regular intervals and when the network topology changes. Each router "advertises" routing information by sending a routing information update every 30 seconds. If a router doesn't receive an update from another router for 180 seconds, those routes provided by that router are declared invalid. The routes are removed from the routing table, but they remain in the RIP routes table. After another 120 seconds without receiving an update for those routes, the routes are removed from regular updates.

When a router receives a routing update that includes changes to an entry, it updates its routing table to reflect the new route. The metric value for the path is increased by 1, and the sender is indicated as the next hop. RIP routers maintain only the best route (the route with the lowest metric value) to a destination.

For more information see The Configuration Menu, Routing Information Protocol Configuration in the *Enterprise NOS Command Reference*.

RIPv1

RIP version 1 uses broadcast User Datagram Protocol (UDP) data packets for the regular routing updates. The main disadvantage is that the routing updates do not carry subnet mask information. Hence, the router cannot determine whether the route is a subnet route or a host route. It is of limited usage after the introduction of RIPv2. For more information about RIPv1 and RIPv2, refer to RFC 1058 and RFC 2453.

RIPv2

RIPv2 is the most popular and preferred configuration for most networks. RIPv2 expands the amount of useful information carried in RIP messages and provides a measure of security. For a detailed explanation of RIPv2, refer to RFC 1723 and RFC 2453.

RIPv2 improves efficiency by using multicast UDP (address 224.0.0.9) data packets for regular routing updates. Subnet mask information is provided in the routing updates. A security option is added for authenticating routing updates, by using a shared password. Enterprise NOS supports using clear password for RIPv2.

RIPv2 in RIPv1 Compatibility Mode

Enterprise NOS allows you to configure RIPv2 in RIPv1compatibility mode, for using both RIPv2 and RIPv1 routers within a network. In this mode, the regular routing updates use broadcast UDP data packet to allow RIPv1 routers to receive those packets. With RIPv1 routers as recipients, the routing updates have to carry natural or host mask. Hence, it is not a recommended configuration for most network topologies.

Note: When using both RIPv1 and RIPv2 within a network, use a single subnet mask throughout the network.

RIP Features

Enterprise NOS provides the following features to support RIPv1 and RIPv2:

Poison Reverse

Simple split horizon in RIP omits routes learned from one neighbor in updates sent to that neighbor. That is the most common configuration used in RIP, with the Poison Reverse feature disabled. Split horizon with poisoned reverse enabled includes such routes in updates, but sets their metrics to 16. The disadvantage of using this feature is the increase of size in the routing updates.

Triggered Updates

Triggered updates are an attempt to speed up convergence. When Triggered Updates is enabled, whenever a router changes the metric for a route, it sends update messages almost immediately, without waiting for the regular update interval. It is recommended to enable Triggered Updates.

Multicast

RIPv2 messages use IPv4 multicast address (224.0.0.9) for periodic updates. Multicast RIPv2 updates are not processed by RIPv1 routers. IGMP is not needed since these are inter-router messages which are not forwarded.

To configure RIPv2 in RIPv1 compatibility mode, set multicast to `disable`, and set version to `both`.

Default Route

The RIP router can listen and supply a default route, usually represented as IPv4 0.0.0.0 in the routing table. When a router does not have an explicit route to a destination network in its routing table, it uses the default route to forward those packets.

Metric

The metric field contains a configurable value between 1 and 15 (inclusive) which specifies the current metric for the interface. The metric value typically indicates the total number of hops to the destination. The metric value of 16 represents an unreachable destination.

Authentication

RIPv2 authentication uses plain text password for authentication. If configured using Authentication password, then it is necessary to enter an authentication key value.

The following method is used to authenticate a RIP message:

- If the router is not configured to authenticate RIPv2 messages, then RIPv1 and unauthenticated RIPv2 messages are accepted; authenticated RIPv2 messages are discarded.
- If the router is configured to authenticate RIPv2 messages, then RIPv1 and RIPv2 messages which pass authentication testing are accepted; unauthenticated and failed authentication RIPv2 messages are discarded.

For maximum security, RIPv1 messages are ignored when authentication is enabled (**`interface ip`** <*x*>**`/ ip rip auth type/password`**); otherwise, the routing information from authenticated messages is propagated by RIPv1 routers in an unauthenticated manner.

RIP Configuration Example

Note: An interface RIP disabled uses all the default values of the RIP, no matter how the RIP parameters are configured for that interface. RIP sends out RIP regular updates to include an UP interface, but not a DOWN interface.

1. Add VLANs for routing interfaces.

```
CN 4093(config)# vlan 2
CN 4093(config-vlan)# exit

CN 4093(config)# interface port 2
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan add 2
CN 4093(config-if)# exit

Port 2 is an UNTAGGED port and its current PVID is 1.
Confirm changing PVID from 1 to 2 [y/n]: y

CN 4093(config)# vlan 3
CN 4093(config-vlan)# exit

CN 4093(config)# interface port 3
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan add 3
CN 4093(config-if)# exit

Port 3 is an UNTAGGED port and its current PVID is 1.
Confirm changing PVID from 1 to 3 [y/n]: y
```

2. Add IP interfaces with IPv4 addresses to VLANs.

```
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# ip address 102.1.1.1
CN 4093(config-ip-if)# vlan 2
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# ip address 103.1.1.1
CN 4093(config-ip-if)# vlan 3
CN 4093(config-ip-if)# exit
```

3. Turn on RIP globally and enable RIP for each interface.

```
CN 4093(config)# router rip
CN 4093(config-router-rip)# enable
CN 4093(config-router-rip)# exit

CN 4093# interface ip 2
CN 4093(config-ip-if)# ip rip enable
CN 4093(config-ip-if)# exit

CN 4093# interface ip 3
CN 4093(config-ip-if)# ip rip enable
CN 4093(config-ip-if)# exit
```

Use the following command to check the current valid routes in the routing table of the switch:

```
CN 4093# show ip route
```

For those RIP learnt routes within the garbage collection period, that are routes phasing out of the routing table with metric 16, use the following command:

```
CN 4093# show ip rip routes
```

Locally configured static routes do not appear in the RIP Routes table.

Chapter 27. Internet Group Management Protocol

Internet Group Management Protocol (IGMP) is used by IPv4 Multicast routers to learn about the existence of host group members on their directly attached subnet (see RFC 2236). The IPv4 Multicast routers get this information by broadcasting IGMP Membership Queries and listening for IPv4 hosts reporting their host group memberships. This process is used to set up a client/server relationship between an IPv4 Multicast source that provides the data streams and the clients that want to receive the data.

The CN4093 10 Gb Converged Scalable Switch (CN4093) can perform IGMP Snooping, or act as an IGMP Relay (proxy) device.

Note: Enterprise NOS 8.4 does not support IPv6 for IGMP.

The following topics are discussed in this chapter:

IGMP Snooping

IGMP Snooping allows the switch to forward multicast traffic only to those ports that request it. IGMP Snooping prevents multicast traffic from being flooded to all ports. The switch learns which server hosts are interested in receiving multicast traffic, and forwards it only to ports connected to those servers.

IGMP Snooping conserves bandwidth. With IGMP Snooping, the switch learns which ports are interested in receiving multicast data, and forwards multicast data only to those ports. In this way, other ports are not burdened with unwanted multicast traffic.

The switch can sense IGMP Membership Reports from attached clients and act as a proxy to set up a dedicated path between the requesting host and a local IPv4 Multicast router. After the pathway is established, the switch blocks the IPv4 Multicast stream from flowing through any port that does not connect to a host member, thus conserving bandwidth.

The client-server path is set up as follows:

- An IPv4 Multicast Router (Mrouter) sends *Membership Queries* to the switch, which forwards them to all ports in a given VLAN.
- Hosts that want to receive the multicast data stream send *Membership Reports* to the switch, which sends a proxy Membership Report to the Mrouter.
- The switch sets up a path between the Mrouter and the host, and blocks all other ports from receiving the multicast.
- Periodically, the Mrouter sends Membership Queries to ensure that the host wants to continue receiving the multicast. If a host fails to respond with a Membership Report, the Mrouter stops sending the multicast to that path.
- The host can send an IGMP Leave packet to the switch, which responds with an IGMP Groups Specific Query in order to check if there are other clients that want to receive the multicast traffic for the group referenced in the Leave packet. If an IGMP Report is not received, the group is deleted from the port and the multicast path is terminated. The switch then sends a Proxy Leave packet to the Mrouter in order to update it. If the FastLeave option is enabled on a VLAN, the multicast path is terminated immediately and the Leave packet is directly forwarded to the Mrouter.

IGMP Groups

The CN4093 supports a maximum of 3072 IGMP entries, on a maximum of 1024 (1022 in stacking mode) VLANs. One IGMP entry is allocated for each unique join request, based on the VLAN and IGMP group address only (regardless of the port). If multiple ports join the same IGMP group using the same VLAN, only a single IGMP entry is used.

IGMPv3

IGMPv3 includes new membership report messages to extend IGMP functionality. The CN4093 provides snooping capability for all types of IGMP version 3 (IGMPv3) Membership Reports, as described in RFC 3376.

IGMPv3 supports Source-Specific Multicast (SSM). SSM identifies session traffic by both source and group addresses. The CN4093 uses *source filtering*, which allows hosts to report interest in receiving multicast packets only from specific source addresses, or from all but specific source addresses.

The CN4093 supports the following IGMPv3 filter modes:

- INCLUDE mode: The host requests membership to a multicast group and provides a list of IPv4 addresses from which it wants to receive traffic.
- EXCLUDE mode: The host requests membership to a multicast group and provides a list of IPv4 addresses from which it *does not* want to receive traffic. This indicates that the host wants to receive traffic only from sources that are not part of the Exclude list. To disable snooping on EXCLUDE mode reports, use the following command:
  `CN 4093(config)# no ip igmp snoop igmpv3 exclude`

By default, the CN4093 snoops the first eight sources listed in the IGMPv3 Group Record. Use the following command to change the number of snooping sources:

```
CN 4093(config)# ip igmp snoop igmpv3 sources <1-64>
```

IGMPv3 Snooping is compatible with IGMPv1 and IGMPv2 Snooping. You can disable snooping on version 1 and version 2 reports, using the following command:

```
CN 4093(config)# no ip igmp snoop igmpv3 v1v2
```

IGMP Snooping Configuration Example

This section provides steps to configure IGMP Snooping on the CN4093, using the Command-Line Interface (CLI).

1. Configure port and VLAN membership on the switch.
2. Add VLANs to IGMP Snooping and enable IGMP Snooping.

```
CN 4093(config)# ip igmp snoop vlan 1
CN 4093(config)# ip igmp snoop enable
```

3. Enable IGMPv3 Snooping (optional).

```
CN 4093(config)# ip igmp snoop igmpv3 enable
```

4. Enable IGMP.

```
CN 4093(config)# ip igmp enable        (Turn on IGMP)
```

5. View dynamic IGMP information.

   To display information about IGMP Groups:

```
CN 4093# show ip igmp groups

Total entries: 5  Total IGMP groups: 2
Note: The <Total IGMP groups> number is computed as
      the number of unique (Group, Vlan) entries!

Note: Local groups (224.0.0.x) are not snooped/relayed and will not appear.
     Source       Group   VLAN   Port  Version  Mode  Expires  Fwd
 ------------ ----------- ---- ------ --------  ----- -------  ---
 10.1.1.1      232.1.1.1     2      4     V3      INC    4:16  Yes
 10.1.1.5      232.1.1.1     2      4     V3      INC    4:16  Yes
    *          232.1.1.1     2      4     V3      INC     -    No
 10.10.10.43   235.0.0.1     9      1     V3      EXC    2:26  No
    *          235.0.0.1     9      1     V3      EXC      -   Yes
```

   To display information about Mrouters learned by the switch:

```
CN 4093# show ip igmp mrouter

Total entries: 3 Total number of dynamic mrouters: 2
Total number of installed static mrouters : 1

SrcIP          VLAN    Port     Version   Expires   MRT       QRV  QQIC
-------------- ------- -------  --------- --------  -------   ----  ----
10.1.1.1          2    EXT18      V3      4:09     128         2    125
10.1.1.5          2    EXT19      V2      4:09     125         -      -
10.10.10.43       9    EXT10      V2     static    -           -      -
```

Note: If IGMP Snooping v1/v2 is enabled and IGMPv3 Snooping is disabled, the output of IGMPv3 reports and queries show some items as IGMPv3 (V3), though they retain v2 behavior. For example, the Source IPv4 address is not relevant for v2 entries.

Static Multicast Router

A static multicast router (Mrouter) can be configured for a particular port on a particular VLAN.

A total of 128 static Mrouters can be configured on the CN4093. Both internal and external ports can accept a static Mrouter.

Note: When static Mrouters are used, the switch will continue learning dynamic Mrouters via IGMP snooping. However, dynamic Mrouters may not replace static Mrouters. If a dynamic Mrouter has the same port and VLAN combination as a static Mrouter, the dynamic Mrouter will not be learned.

Following is an example of configuring a static multicast router:

1. For each Mrouter, configure a port, VLAN, and IGMP version of the multicast router.

```
CN 4093(config)# ip igmp mrouter EXT5 1 2
```

2. Verify the configuration.

```
CN 4093(config)# show ip igmp mrouter
```

IGMP Relay

The CN4093 can act as an IGMP Relay (or IGMP Proxy) device that relays IGMP multicast messages and traffic between an Mrouter and end stations. IGMP Relay allows the CN4093 to participate in network multicasts with no configuration of the various multicast routing protocols, so you can deploy it in the network with minimal effort.

To an IGMP host connected to the CN4093, IGMP Relay appears to be an IGMP multicast router (Mrouter). IGMP Relay sends Membership Queries to hosts, which respond by sending an IGMP response message. A host can also send an unsolicited Join message to the IGMP Relay.

To a multicast router, IGMP Relay appears as a host. The Mrouter sends IGMP host queries to IGMP Relay, and IGMP Relay responds by forwarding IGMP host reports and unsolicited join messages from its attached hosts.

IGMP Relay also forwards multicast traffic between the Mrouter and end stations, similar to IGMP Snooping.

You can configure up to two Mrouters to use with IGMP Relay. One Mrouter acts as the primary Mrouter, and one is the backup Mrouter. The CN4093 uses ICMP health checks to determine if the primary and backup mrouters are reachable.

Configuration Guidelines

Consider the following guidelines when you configure IGMP Relay:

- IGMP Relay is supported in stand-alone (non-stacking) mode only.
- IGMP Relay and IGMP Snooping/Querier are mutually exclusive—if you enable IGMP Relay, you must turn off IGMP Snooping/Querier.
- Add VLANs to the IGMP Relay list, using the following command:

```
CN 4093(config)# ip igmp relay vlan <VLAN ID>
```

- If IGMP hosts reside on different VLANs, you must:
  - Disable IGMP flooding.

```
CN 4093(config)# vlan <VLAN ID>
CN 4093(config-vlan)# no flood
```

  - Enable CPU forwarding to ensure that multicast data is forwarded across the VLANs.

```
CN 4093(config)# vlan <VLAN ID>
CN 4093(config-vlan)# cpu
```

IGMP Relay Configuration Example

Use the following procedure to configure IGMP Relay.

1. Configure IP interfaces with IPv4 addresses and assign VLANs.

```
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 10.10.1.1 255.255.255.0 enable
CN 4093(config-ip-if)# vlan 2
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ip address 10.10.2.1 255.255.255.0 enable
CN 4093(config-ip-if)# vlan 3
CN 4093(config-ip-if)# exit
```

2. Turn IGMP on.

```
CN 4093(config)# ip igmp enable
```

3. Configure the upstream Mrouters with IPv4 addresses.

```
CN 4093(config)# ip igmp relay mrouter 1 address 100.0.1.2
CN 4093(config)# ip igmp relay mrouter 1 enable
CN 4093(config)# ip igmp relay mrouter 2 address 100.0.2.4
CN 4093(config)# ip igmp relay mrouter 2 enable
```

4. Add VLANs to the downstream network and enable IGMP Relay

```
CN 4093(config)# ip igmp relay vlan 2
CN 4093(config)# ip igmp relay vlan 3
CN 4093(config)# ip igmp relay enable
```

IGMP Querier

IGMP Querier allows the switch to perform the multicast router (Mrouter) role and provide Mrouter discovery when the network or virtual LAN (VLAN) does not have a router.

When the IGMP Querier feature is enabled on a VLAN, the switch participates in the Querier election process and has the possibility to be elected as Querier for the VLAN. The IGMP querier periodically broadcasts IGMP Queries and listens for hosts to respond with IGMP Reports indicating their IGMP group memberships. If multiple Mrouters exist on a given network, the Mrouters elect one as the querier, which performs all periodic membership queries. The election process can be based on IPv4 address or MAC address.

Note: When IGMP Querier is enabled on a VLAN, the switch performs the role of IGMP querier only if it meets the IGMP querier election criteria.

IGMP Querier Configuration Example

Follow this procedure to configure IGMP Querier.

1. Enable IGMP and configure the source IPv4 address for IGMP Querier on a VLAN.

```
CN 4093(config)# ip igmp enable
CN 4093(config)# ip igmp querier vlan 2 source-ip 10.10.10.1
```

2. Enable IGMP Querier on the VLAN.

```
CN 4093(config)# ip igmp querier vlan 2 enable
```

3. Configure the querier election type and define the address.

```
CN 4093(config)# ip igmp querier vlan 2 election-type ipv4
```

4. Verify the configuration.

```
CN 4093# show ip igmp querier vlan 2
Current VLAN 2 IGMP querier settings: ON
        querier type: ipv4
        max response time: 100
        querier interval: 125
        Querier robustness: 2
        source IP: 10.10.10.15
        startup count: 2
        startup query interval: 31
        version: v3
```

Additional IGMP Features

The following topics are discussed in this section:

FastLeave

In normal IGMP operation, when the switch receives an IGMPv2 *leave* message, it sends a Group-Specific Query to determine if any other devices in the same group (and on the same port) are still interested in the specified multicast group traffic. The switch removes the affiliated port from that particular group, if it does not receive an IGMP Membership Report within the query-response-interval.

With FastLeave enabled on the VLAN, a port can be removed immediately from the port list of the group entry when the IGMP Leave message is received, unless a multicast router was learned on the port.

Enable FastLeave only on VLANs that have only one host connected to each physical port.

IGMP Filtering

With IGMP Filtering, you can allow or deny a port to learn certain IGMP or IPMC groups. This allows you to restrict users from receiving certain multicast traffic.

If access to a multicast group is denied, IGMP Membership Reports from the port are dropped, and the port is not allowed to receive IPv4 multicast traffic from that group. If access to the multicast group is allowed, Membership Reports from the port are forwarded for normal processing.

To configure IGMP Filtering, you must globally enable IGMP filtering, define an IGMP filter, assign the filter to a port, and enable IGMP Filtering on the port. To define an IGMP filter, you must configure a range of IPv4 multicast groups, choose whether the filter will allow or deny multicast traffic for groups within the range, and enable the filter.

Configuring the Range

Each IGMP Filter allows you to set a start and end point that defines the range of IPv4 addresses upon which the filter takes action. Each IPv4 address in the range must be between 224.0.0.0 and 239.255.255.255.

Configuring the Action

Each IGMP filter can allow or deny IPv4 multicasts to the range of IPv4 addresses configured. If you configure the filter to deny IPv4 multicasts, then IGMP Membership Reports from multicast groups within the range are dropped. You can configure a secondary filter to allow IPv4 multicasts to a small range of addresses within a larger range that a primary filter is configured to deny. The two filters work together to allow IPv4 multicasts to a small subset of addresses within the larger range of addresses.

Note: Lower-numbered filters take precedence over higher-number filters. For example, the action defined for IGMP Filter 1 supersedes the action defined for IGMP Filter 2.

IGMP Filtering Configuration Example

1. Enable IGMP filtering on the switch.

```
CN 4093(config)# ip igmp filtering
```

2. Define an IGMP filter with IPv4 information.

```
CN 4093(config)# ip igmp profile 1 range 225.0.0.0 226.0.0.0
CN 4093(config)# ip igmp profile 1 action deny
CN 4093(config)# ip igmp profile 1 enable
```

3. Assign the IGMP filter to a port.

```
CN 4093(config)# interface port 3
CN 4093(config-if)# ip igmp profile 1
CN 4093(config-if)# ip igmp filtering
```

Chapter 28. Multicast Listener Discovery

Multicast Listener Discovery (MLD) is an IPv6 protocol that a host uses to request multicast data for a multicast group. An IPv6 router uses MLD to discover the presence of multicast listeners (nodes that want to receive multicast packets) on its directly attached links, and to discover specifically the multicast addresses that are of interest to those neighboring nodes.

MLD version 1 is derived from Internet Group Management Protocol version 2 (IGMPv2) and MLDv2 is derived from IGMPv3. MLD uses ICMPv6 (IP Protocol 58) message types. See RFC 2710 and RFC 3810 for details.

MLDv2 protocol, when compared to MLDv1, adds support for source filtering—the ability for a node to report interest in listening to packets only from specific source addresses, or from all but specific source addresses, sent to a particular multicast address. MLDv2 is interoperable with MLDv1. See RFC 3569 for details on Source-Specific Multicast (SSM).

The following topics are discussed in this chapter:

MLD Terms

Following are the commonly used MLD terms:

- Multicast traffic: Flow of data from one source to multiple destinations.
- Group: A multicast stream to which a host can join.
- Multicast Router (Mrouter): A router configured to make routing decisions for multicast traffic. The router identifies the type of packet received (unicast or multicast) and forwards the packet to the intended destination.
- Querier: An Mrouter that sends periodic query messages. Only one Mrouter on the subnet can be elected as the Querier.
- Multicast Listener Query: Messages sent by the Querier. There are three types of queries:
  - General Query: Sent periodically to learn multicast address listeners from an attached link. CN4093 uses these queries to build and refresh the Multicast Address Listener state. General Queries are sent to the link-scope all-nodes multicast address (FF02::1), with a multicast address field of 0, and a maximum response delay of *query response interval*.
  - Multicast Address Specific Query: Sent to learn if a specific multicast address has any listeners on an attached link. The multicast address field is set to the IPv6 multicast address.
  - Multicast Address and Source Specific Query: Sent to learn if, for a specified multicast address, there are nodes still listening to a specific set of sources. Supported only in MLDv2.

  Note: Multicast Address Specific Queries and Multicast Address and Source Specific Queries are sent only in response to State Change Reports, and never in response to Current State Reports.
- Multicast Listener Report: Sent by a host when it joins a multicast group, or in response to a Multicast Listener Query sent by the Querier. Hosts use these reports to indicate their current multicast listening state, or changes in the multicast listening state of their interfaces. These reports are of two types:
  - Current State Report: Contains the current Multicast Address Listening State of the host.
  - State Change Report: If the listening state of a host changes, the host immediately reports these changes through a State Change Report message. These reports contain either Filter Mode Change records and/or Source List Change records. State Change Reports are retransmitted several times to ensure all Mrouters receive it.
- Multicast Listener Done: Sent by a host when it wants to leave a multicast group. This message is sent to the link-scope all-routers IPv6 destination address of FF02::2. When an Mrouter receives a Multicast Listener Done message from the last member of the multicast address on a link, it stops forwarding traffic to this multicast address.

How MLD Works

The software uses the information obtained through MLD to maintain a list of multicast group memberships for each interface and forwards the multicast traffic only to interested listeners.

Without MLD, the switch forwards IPv6 multicast traffic through all ports, increasing network load. Following is an overview of operations when MLD is configured on CN4093:

- The switch acts as an Mrouter when MLDv1/v2 is configured and enabled on each of its directly attached links. If the switch has multiple interfaces connected to the same link, it operates the protocol on any one of the interfaces.
- If there are multiple Mrouters on the subnet, the Mrouter with the numerically lowest IPv6 address is elected as the Querier.
- The Querier sends general queries at short intervals to learn multicast address listener information from an attached link.
- Hosts respond to these queries by reporting their per-interface Multicast Address Listening state, through Current State Report messages sent to a specific multicast address that all MLD routers on the link listen to.
- If the listening state of a host changes, the host immediately reports these changes through a State Change Report message.
- The Querier sends a Multicast Address Specific Query to verify if hosts are listening to a specified multicast address or not. Similarly, if MLDv2 is configured, the Querier sends a Multicast Address and Source Specific Query to verify, for a specified multicast address, if hosts are listening to a specific set of sources, or not. MLDv2 listener report messages consists of Multicast Address Records:
  - INCLUDE: to receive packets from source specified in the MLDv2 message
  - EXCLUDE: to receive packets from all sources except the ones specified in the MLDv2 message
- A host can send a State Change Report to indicate its desire to stop listening to a particular multicast address (or source in MLDv2). The Querier then sends a multicast address specific query to verify if there are other listeners of the multicast address. If there aren't any, the Mrouter deletes the multicast address from its Multicast Address Listener state and stops sending multicast traffic. Similarly in MLDv2, the Mrouter sends a Multicast Address and Source Specific Query to verify if, for a specified multicast address, there are hosts still listening to a specific set of sources.

CN4093 supports MLD versions 1 and 2.

Note: MLDv2 operates in version 1 compatibility mode when, in a specific network, not all hosts are configured with MLDv2.

How Flooding Impacts MLD

When `flood` option is disabled, the unknown multicast traffic is discarded if no Mrouters are learned on the switch. You can set the flooding behavior by configuring the `flood` and `cpu` options. You can optimize the flooding to ensure that unknown IP multicast (IPMC) data packets are not dropped during the learning phase.

The flooding options include:

- `flood`: Enable hardware flooding in VLAN for the unregistered IPMC; This option is enabled by default.
- `cpu`: Enable sending unregistered IPMC to the Mrouter ports. However, during the learning period, there will be some packet loss. The `cpu` option is enabled by default. You must ensure that the `flood` and `optflood` options are disabled.
- `optflood`: Enable optimized flooding to allow sending the unregistered IPMC to the Mrouter ports without having any packet loss during the learning period; This option is disabled by default; When `optflood` is enabled, the `flood` and `cpu` settings are ignored.

The flooding parameters must be configured per VLAN. Enter the following command to set the `flood` or `cpu` options:

```
CN 4093(config)# vlan <VLAN ID>
CN 4093(config-vlan)# [no] flood
CN 4093(config-vlan)# [no] cpu
CN 4093(config-vlan)# [no] optflood
```

MLD Querier

An Mrouter acts as a Querier and periodically (at short query intervals) sends query messages in the subnet. If there are multiple Mrouters in the subnet, only one can be the Querier. All Mrouters on the subnet listen to the messages sent by the multicast address listeners, and maintain the same multicast listening information state.

All MLDv2 queries are sent with the FE80::/64 link-local source address prefix.

Querier Election

Only one Mrouter can be the Querier per subnet. All other Mrouters will be non-Queriers. MLD versions 1 and 2 elect the Mrouter with the numerically lowest IPv6 address as the Querier.

If the switch is configured as an Mrouter on a subnet, it also acts as a Querier by default and sends multiple general queries. If the switch receives a general query from another Querier with a numerically lower IPv6 address, it sets the *other querier present timer* to the *other querier present timeout*, and changes its state to non-Querier. When the *other querier present timer* expires, it regains the Querier state and starts sending general queries.

Note: When MLD Querier is enabled on a VLAN, the switch performs the role of an MLD Querier only if it meets the MLD Querier election criteria.

Dynamic Mrouters

The switch learns Mrouters on the ingress VLANs of the MLD-enabled interface. All report or done messages are forwarded to these Mrouters. By default, the option of dynamically learning Mrouters is disabled. To enable it, use the following command:

```
CN 4093(config)# interface ip <interface number>
CN 4093(config-ip-if)# ipv6 mld dmrtr enable
```

MLD Capacity and Default Values

Table 36 lists the maximum and minimum values of the CN4093 variables.

Table 36. *CN4093 Capacity Table*

Variable	Maximum Value
IPv6 Multicast Entries	256
IPv6 Interfaces for MLD	8

Table 37 lists the default settings for MLD features and variables.

Table 37. *MLD Timers and Default Values*

Field	Default Value
Robustness Variable (RV)	2
Query Interval (QI)	125 seconds
Query Response Interval (QRI)	10 seconds
Multicast Address Listeners Interval (MALI)	260 seconds [derived: RV*QI+QRI]
Other Querier Present Interval [OQPT]	255 seconds [derived: RV*QI + ½ QRI]
Start up Query Interval [SQI]	31.25 seconds [derived: ¼ * QI]
Startup Query Count [SQC]	2 [derived: RV]
Last Listener Query Interval [LLQI]	1 second
Last Listener Query Count [LLQC]	2 [derived: RV]
Last Listener Query Time [LLQT]	2 seconds [derived: LLQI * LLQT]
Older Version Querier Present Timeout: [OVQPT]	260 seconds [derived: RV*QI+ QRI]
Older Version Host Present Interval [OVHPT]	260 seconds [derived: RV* QI+QRI]

Configuring MLD

Following are the steps to enable MLD and configure the interface parameters:

1. Turn on MLD globally.

```
CN 4093(config)# ipv6 mld
CN 4093(config-router-mld)# enable
CN 4093(config-router-mld)# exit
```

2. Create an IPv6 interface.

```
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# ipv6 address 2002:1:0:0:0:0:0:3
CN 4093(config-ip-if)# ipv6 prefixlen 64
```

3. Enable MLD on the IPv6 interface.

```
CN 4093(config-ip-if)# ipv6 mld enable
```

4. Configure the MLD parameters on the interface: version, robustness, query response interval, MLD query interval, and last listener query interval.

```
CN 4093(config-ip-if)# ipv6 mld version <1-2>       (MLD version)
CN 4093(config-ip-if)# ipv6 mld robust <1-10>       (Robustness)
CN 4093(config-ip-if)# ipv6 mld qri <1-256>         (In seconds)
CN 4093(config-ip-if)# ipv6 mld qintrval <1-608>    (In seconds)
CN 4093(config-ip-if)# ipv6 mld llistnr <1-32>      (In seconds)
```

Chapter 29. Border Gateway Protocol

Border Gateway Protocol (BGP) is an Internet protocol that enables routers on an IPv4 network to share and advertise routing information with each other about the segments of the IPv4 address space they can access within their network and with routers on external networks. BGP allows you to decide what is the "best" route for a packet to take from your network to a destination on another network rather than simply setting a default route from your border router(s) to your upstream provider(s). BGP is defined in RFC 1771.

CN4093 10 Gb Converged Scalable Switches (CN4093s) can advertise their IP interfaces and IPv4 addresses using BGP and take BGP feeds from as many as BGP router peers. This allows more resilience and flexibility in balancing traffic from the Internet.

Note: Enterprise NOS 8.4 does not support IPv6 for BGP.

The following topics are discussed in this section:

Internal Routing Versus External Routing

To ensure effective processing of network traffic, every router on your network needs to know how to send a packet (directly or indirectly) to any other location/destination in your network. This is referred to as *internal routing* and can be done with static routes or using active, internal dynamic routing protocols, such as RIP, RIPv2, and OSPF.

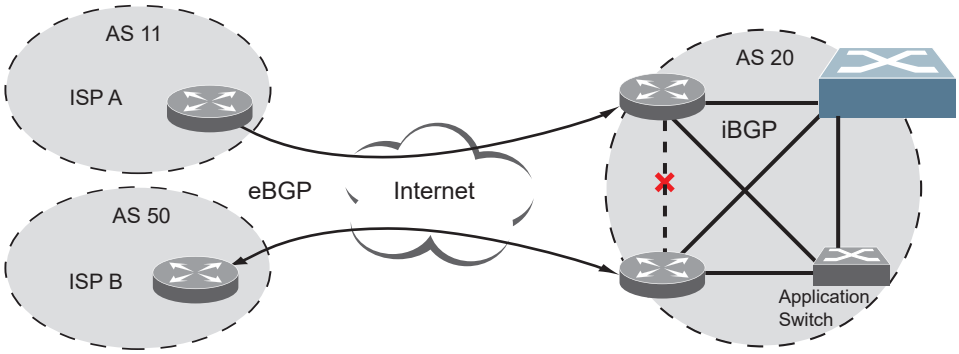
Static routes should have a higher degree of precedence than dynamic routing protocols. If the destination route is not in the route cache, then the packets are forwarded to the default gateway which may be incorrect if a dynamic routing protocol is enabled.

It is also useful to tell routers outside your network (upstream providers or *peers*) about the routes you can access in your network. External networks (those outside your own) that are under the same administrative control are referred to as *autonomous systems* (AS). Sharing of routing information between autonomous systems is known as *external routing*.

External BGP (eBGP) is used to exchange routes between different autonomous systems whereas internal BGP (iBGP) is used to exchange routes within the same autonomous system. An iBGP is a type of internal routing protocol you can use to do active routing inside your network. It also carries AS path information, which is important when you are an ISP or doing BGP transit.

The iBGP peers have to maintain reciprocal sessions to every other iBGP router in the same AS (in a full-mesh manner) in order to propagate route information throughout the AS. If the iBGP session shown between the two routers in AS 20 was not present (as indicated in Figure 43), the top router would not learn the route to AS 50, and the bottom router would not learn the route to AS 11, even though the two AS 20 routers are connected via the Flex System and the Application Switch.

Figure 43. iBGP and eBGP



Typically, an AS has one or more *border routers*—peer routers that exchange routes with other ASs—and an internal routing scheme that enables routers in that AS to reach every other router and destination within that AS. When you *advertise* routes to border routers on other autonomous systems, you are effectively committing to carry data to the IPv4 space represented in the route being advertised. For example, if you advertise 192.204.4.0/24, you are declaring that if another router sends you data destined for any address in 192.204.4.0/24, you know how to carry that data to its destination.

Forming BGP Peer Routers

Two BGP routers become peers or neighbors once you establish a TCP connection between them. For each new route, if a peer is interested in that route (for example, if a peer would like to receive your static routes and the new route is static), an update message is sent to that peer containing the new route. For each route removed from the route table, if the route has already been sent to a peer, an update message containing the route to withdraw is sent to that peer.

For each Internet host, you must be able to send a packet to that host, and that host has to have a path back to you. This means that whoever provides Internet connectivity to that host must have a path to you. Ultimately, this means that they must “hear a route” which covers the section of the IPv4 space you are using; otherwise, you will not have connectivity to the host in question.

What is a Route Map?

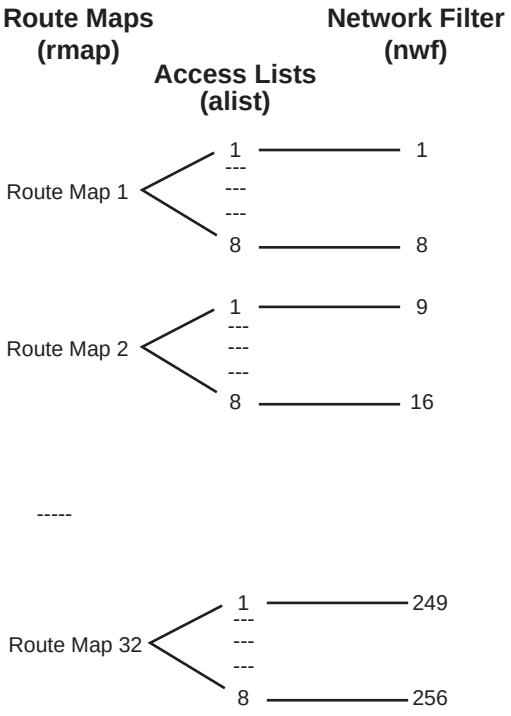
A route map is used to control and modify routing information. Route maps define conditions for redistributing routes from one routing protocol to another or controlling routing information when injecting it in and out of BGP. For example, a route map is used to set a preference value for a specific route from a peer router and another preference value for all other routes learned via the same peer router. For example, the following commands are used to define a route map:

```
CN 4093(config)# route-map <map number>      (Select a route map)
CN 4093(config-route-map)# ?                 (List available commands)
```

A route map allows you to match attributes, such as metric, network address, and AS number. It also allows users to overwrite the local preference metric and to append the AS number in the AS route. See "BGP Failover Configuration" on page 466.

Enterprise NOS allows you to configure 32 route maps. Each route map can have up to eight access lists. Each access list consists of a network filter. A network filter defines an IPv4 address and subnet mask of the network that you want to include in the filter. Figure 44 illustrates the relationship between route maps, access lists and network filters.

Figure 44. Distributing Network Filters in Access Lists and Route Maps



Incoming and Outgoing Route Maps

You can have two types of route maps: incoming and outgoing. A BGP peer router can be configured to support up to eight route maps in the incoming route map list and outgoing route map list.

If a route map is not configured in the incoming route map list, the router imports all BGP updates. If a route map is configured in the incoming route map list, the router ignores all unmatched incoming updates. If you set the action to **deny**, you must add another route map to permit all unmatched updates.

Route maps in an outgoing route map list behave similar to route maps in an incoming route map list. If a route map is not configured in the outgoing route map list, all routes are advertised or permitted. If a route map in the outgoing route map list is set to **permit**, matched routes are advertised and unmatched routes are ignored.

Precedence

You can set a priority to a route map by specifying a precedence value with the following commands:

```
CN 4093(config)# route-map <map number>              (Select a route map)
CN 4093(config-route-map)# precedence <1-255>   (Specify a precedence)
CN 4093(config-route-map)# exit
```

The smaller the value the higher the precedence. If two route maps have the same precedence value, the smaller number has higher precedence.

Configuration Example

To configure route maps, you need to do the following:

1. Define network filter.

```
CN 4093(config)# ip match-address 1 <IPv4 address> <IPv4 subnet mask>
CN 4093(config)# ip match-address 1 enable
```

Enter a filter number from 1 to 256. Specify the IPv4 address and subnet mask of the network that you want to match. Enable the network filter. You can distribute up to 256 network filters among 32 route maps each containing eight access lists.

2. (Optional) Define the criteria for the access list and enable it.

Specify the access list and associate the network filter number configured in Step 1.

```
CN 4093(config)# route-map 1
CN 4093(config-route-map)# access-list 1 match-address 1
CN 4093(config-route-map)# access-list 1 metric <metric value>
CN 4093(config-route-map)# access-list 1 action deny
CN 4093(config-route-map)# access-list 1 enable
```

Steps 2 and 3 are optional, depending on the criteria that you want to match. In Step 2, the network filter number is used to match the subnets defined in the network filter. In Step 3, the autonomous system number is used to match the subnets. Or, you can use both (Step 2 and Step 3) criteria: access list (network filter) and access path (AS filter) to configure the route maps.

3. (Optional) Configure the attributes in the AS filter menu.

```
CN 4093(config-route-map)# as-path-list 1 as 1
CN 4093(config-route-map)# as-path-list 1 action deny
CN 4093(config-route-map)# as-path-list 1 enable
```

4. Set up the BGP attributes.

   If you want to overwrite the attributes that the peer router is sending, then define the following BGP attributes:

   - Specify up to 32 AS numbers that you want to prepend to a matched route and the local preference for the matched route.
   - Specify the metric [Multi Exit Discriminator (MED)] for the matched route.

```
CN 4093(config-route-map)# as-path-preference <AS number> [<AS number>] ...
CN 4093(config-route-map)# local-preference <local preference value>
CN 4093(config-route-map)# metric <metric value>
```

5. Enable the route map.

```
CN 4093(config-route-map)# enable
CN 4093(config-route-map)# exit
```

6. Turn BGP on.

```
CN 4093(config)# router bgp
CN 4093(config-router-bgp)# enable
```

7. Assign the route map to a peer router.

   Select the peer router and then add the route map to the incoming route map list,

```
CN 4093(config-router-bgp)# neighbor 1 route-map in <1-32>
```

   or to the outgoing route map list.

```
CN 4093(config-router-bgp)# neighbor 1 route-map out <1-32>
```

8. Exit Router BGP mode.

```
CN 4093(config-router-bgp)# exit
```

Aggregating Routes

Aggregation is the process of combining several different routes in such a way that a single route can be advertised, which minimizes the size of the routing table. You can configure aggregate routes in BGP either by redistributing an aggregate route into BGP or by creating an aggregate entry in the BGP routing table.

To define an aggregate route in the BGP routing table, use the following commands:

```
CN 4093(config)# router bgp
CN 4093(config-router-bgp)# aggregate-address <1-16> <IPv4 address> <mask>
CN 4093(config-router-bgp)# aggregate-address <1-16> enable
CN 4093(config-router-bgp)# exit
```

An example of creating a BGP aggregate route is shown in "Default Redistribution and Route Aggregation Example" on page 468.

Redistributing Routes

In addition to running multiple routing protocols simultaneously, Enterprise NOS software can redistribute information from one routing protocol to another. For example, you can instruct the switch to use BGP to re-advertise static routes. This applies to all of the IP-based routing protocols.

You can also conditionally control the redistribution of routes between routing domains by defining a method known as route maps between the two domains. For more information on route maps, see "What is a Route Map?" on page 460. Redistributing routes is another way of providing policy control over whether to export OSPF routes, fixed routes, and static routes. For an example configuration, see "Default Redistribution and Route Aggregation Example" on page 468.

Default routes can be configured using the following methods:

- Import
- Originate—The router sends a default route to peers if it does not have any default routes in its routing table.
- Redistribute—Default routes are either configured through the default gateway or learned via other protocols and redistributed to peer routers. If the default routes are from the default gateway, enable the static routes because default routes from the default gateway are static routes. Similarly, if the routes are learned from another routing protocol, make sure you enable that protocol for redistribution.
- None

BGP Attributes

The following two BGP attributes are discussed in this section: Local preference and metric (Multi-Exit Discriminator).

Local Preference Attribute

When there are multiple paths to the same destination, the local preference attribute indicates the preferred path. The path with the higher preference is preferred (the default value of the local preference attribute is 100). Unlike the weight attribute, which is only relevant to the local router, the local preference attribute is part of the routing update and is exchanged among routers in the same AS.

The local preference attribute can be set in one of two ways:

- Using the BGP default local preference method, affecting the outbound direction only.

```
CN 4093(config)# router bgp
CN 4093(config_router_bgp)# local-preference <0-4294967294>
CN 4093(config_router_bgp)# exit
```

- Using the route map local preference method, which affects both inbound and outbound directions.

```
CN 4093(config)# route-map 1
CN 4093(config_route_map)# local-preference <0-4294967294>
CN 4093(config_route_map)# enabled
CN 4093(config_router_map)# exit
CN 4093(config)# router bgp
CN 4093(config_router_bgp)# neighbor {<number>/group <number>} route-map
{<in/out> <1-255>}
```

Metric (Multi-Exit Discriminator) Attribute

This attribute is a hint to external neighbors about the preferred path into an AS when there are multiple entry points. A lower metric value is preferred over a higher metric value. The default value of the metric attribute is 0.

Unlike local preference, the metric attribute is exchanged between ASs; however, a metric attribute that comes into an AS does not leave the AS.

When an update enters the AS with a certain metric value, that value is used for decision making within the AS. When BGP sends that update to another AS, the metric is reset to 0.

Unless otherwise specified, the router compares metric attributes for paths from external neighbors that are in the same AS.

Selecting Route Paths in BGP

BGP selects only one path as the best path. It does not rely on metric attributes to determine the best path. When the same network is learned via more than one BGP peer, BGP uses its policy for selecting the best route to that network. The BGP implementation on the CN4093 uses the following criteria to select a path when the same route is received from multiple peers.

1. Local fixed and static routes are preferred over learned routes.
2. With iBGP peers, routes with higher local preference values are selected.
3. In the case of multiple routes of equal preference, the route with lower AS path weight is selected.

   AS path weight = 128 **x** AS path length (number of autonomous systems traversed).
4. In the case of equal weight and routes learned from peers that reside in the same AS, the lower metric is selected.

   Note: A route with a metric is preferred over a route without a metric.
5. The lower cost to the next hop of routes is selected.
6. In the case of equal cost, the eBGP route is preferred over iBGP.
7. If all routes have same route type (eBGP or iBGP), the route with the lower router ID is selected.

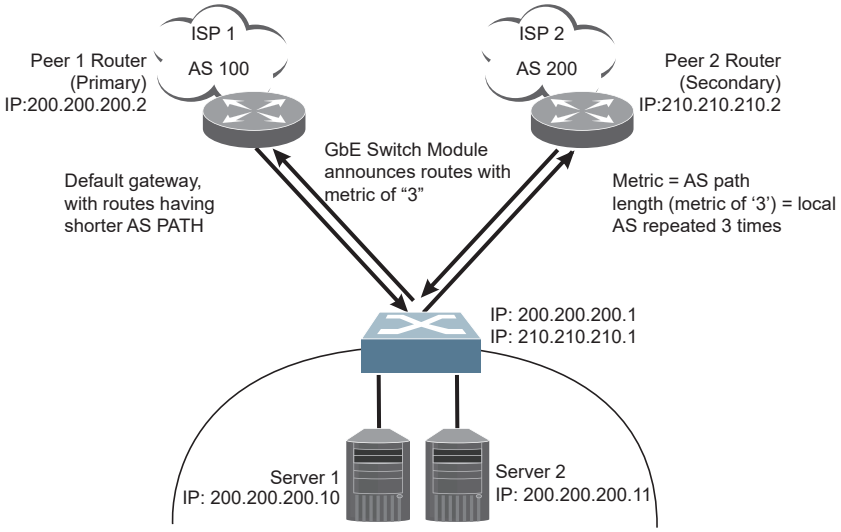
   When the path is selected, BGP puts the selected path in its routing table and propagates the path to its neighbors.

BGP Failover Configuration

Use the following example to create redundant default gateways for a CN4093 at a Web Host/ISP site, eliminating the possibility, should one gateway go down, that requests will be forwarded to an upstream router unknown to the switch.

As shown in Figure 45, the switch is connected to ISP 1 and ISP 2. The customer negotiates with both ISPs to allow the switch to use their peer routers as default gateways. The ISP peer routers will then need to announce themselves as default gateways to the CN4093.

Figure 45. BGP Failover Configuration Example



On the CN4093, one peer router (the secondary one) is configured with a longer AS path than the other, so that the peer with the shorter AS path will be seen by the switch as the primary default gateway. ISP 2, the secondary peer, is configured with a metric of "3," thereby appearing to the switch to be three router *hops* away.

1. Define the VLANs.

   For simplicity, both default gateways are configured in the same VLAN in this example. The gateways could be in the same VLAN or different VLANs.

```
CN 4093(config)# vlan 1
```

2. Define the IP interfaces with IPv4 addresses.

   The switch will need an IP interface for each default gateway to which it will be connected. Each interface must be placed in the appropriate VLAN. These interfaces will be used as the primary and secondary default gateways for the switch.

```
CN 4093(config)# interface ip 1 address 200.200.200.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 2 address 210.210.210.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
```

3. Enable IP forwarding.

   IP forwarding is turned on by default and is used for VLAN-to-VLAN (non-BGP) routing. Make sure IP forwarding is on if the default gateways are on different subnets or if the switch is connected to different subnets and those subnets need to communicate through the switch (which they almost always do).

```
CN 4093(config)# ip routing              (Enable IP forwarding)
```

   Note: To help eliminate the possibility for a Denial of Service (DoS) attack, the forwarding of directed broadcasts is disabled by default.

4. Configure BGP peer router 1 and 2.

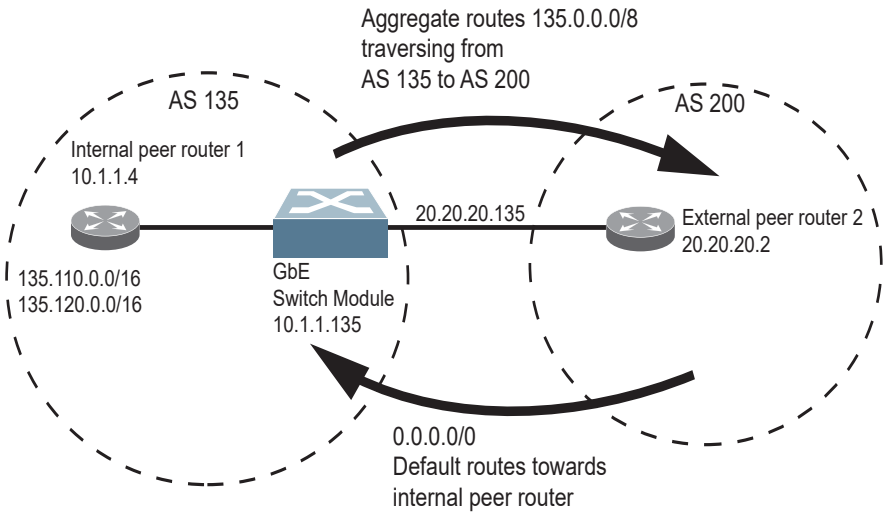
```
CN 4093(config)# router bgp
CN 4093(config-router-bgp)# ip router-id 8.8.8.8
CN 4093(config-router-bgp)# as 816
CN 4093(config-router-bgp)# neighbor 1 remote-address 200.200.200.2
CN 4093(config-router-bgp)# neighbor 1 remote-as 100
CN 4093(config-router-bgp)# no neighbor 1 shutdown
CN 4093(config-router-bgp)# neighbor 2 remote-address 210.210.210.2
CN 4093(config-router-bgp)# neighbor 2 remote-as 200
CN 4093(config-router-bgp)# no neighbor 2 shutdown
```

Default Redistribution and Route Aggregation Example

This example shows you how to configure the switch to redistribute information from one routing protocol to another and create an aggregate route entry in the BGP routing table to minimize the size of the routing table.

As illustrated in Figure 46, you have two peer routers: an internal and an external peer router. Configure the CN4093 to redistribute the default routes from AS 200 to AS 135. At the same time, configure for route aggregation to allow you to condense the number of routes traversing from AS 135 to AS 200.

Figure 46. Route Aggregation and Default Route Redistribution



1. Configure the IP interface.
2. Configure the AS number (AS 135) and router ID number (10.1.1.135).

```
CN 4093(config)# router bgp
CN 4093(config-router-bgp)# as 135
CN 4093(config-router-bgp)# exit
CN 4093(config)# ip router-id 10.1.1.135
```

3. Configure internal peer router 1 and external peer router 2.

```
CN 4093(config)# router bgp
CN 4093(config-router-bgp)# neighbor 1 remote-address 10.1.1.4
CN 4093(config-router-bgp)# neighbor 1 remote-as 135
CN 4093(config-router-bgp)# no neighbor 1 shutdown
CN 4093(config-router-bgp)# neighbor 2 remote-address 20.20.20.2
CN 4093(config-router-bgp)# neighbor 2 remote-as 200
CN 4093(config-router-bgp)# no neighbor 2 shutdown
```

4. Configure redistribution for Peer 1.

```
CN 4093(config-router-bgp)# neighbor 1 redistribute default-action
redistribute
CN 4093(config-router-bgp)# neighbor 1 redistribute fixed
```

5. Configure aggregation policy control.

   Configure the routes that you want aggregated.

```
CN 4093(config-router-bgp)# aggregate-address 1 135.0.0.0 255.0.0.0
CN 4093(config-router-bgp)# aggregate-address 1 enable
```

Chapter 30. OSPF

Enterprise NOS supports the Open Shortest Path First (OSPF) routing protocol. The Enterprise NOS implementation conforms to the OSPF version 2 specifications detailed in Internet RFC 1583, and OSPF version 3 specifications in RFC 5340. The following sections discuss OSPF support for the CN4093 10 Gb Converged Scalable Switch (CN4093):

- "OSPFv2 Overview" on page 471. This section provides information on OSPFv2 concepts, such as types of OSPF areas, types of routing devices, neighbors, adjacencies, link state database, authentication, and internal versus external routing.
- "OSPFv2 Implementation in Enterprise NOS" on page 476. This section describes how OSPFv2 is implemented in Enterprise NOS, such as configuration parameters, electing the designated router, summarizing routes, defining route maps and so forth.
- "OSPFv2 Configuration Examples" on page 487. This section provides step-by-step instructions on configuring different OSPFv2 examples:
  - Creating a simple OSPF domain
  - Creating virtual links
  - Summarizing routes
- "OSPFv3 Implementation in Enterprise NOS" on page 495. This section describes differences and additional features found in OSPFv3.

OSPFv2 Overview

OSPF is designed for routing traffic within a single IP domain called an Autonomous System (AS). The AS can be divided into smaller logical units known as *areas*.

All routing devices maintain link information in their own Link State Database (LSDB). The LSDB for all routing devices within an area is identical but is not exchanged between different areas. Only routing updates are exchanged between areas, thereby significantly reducing the overhead for maintaining routing information on a large, dynamic network.

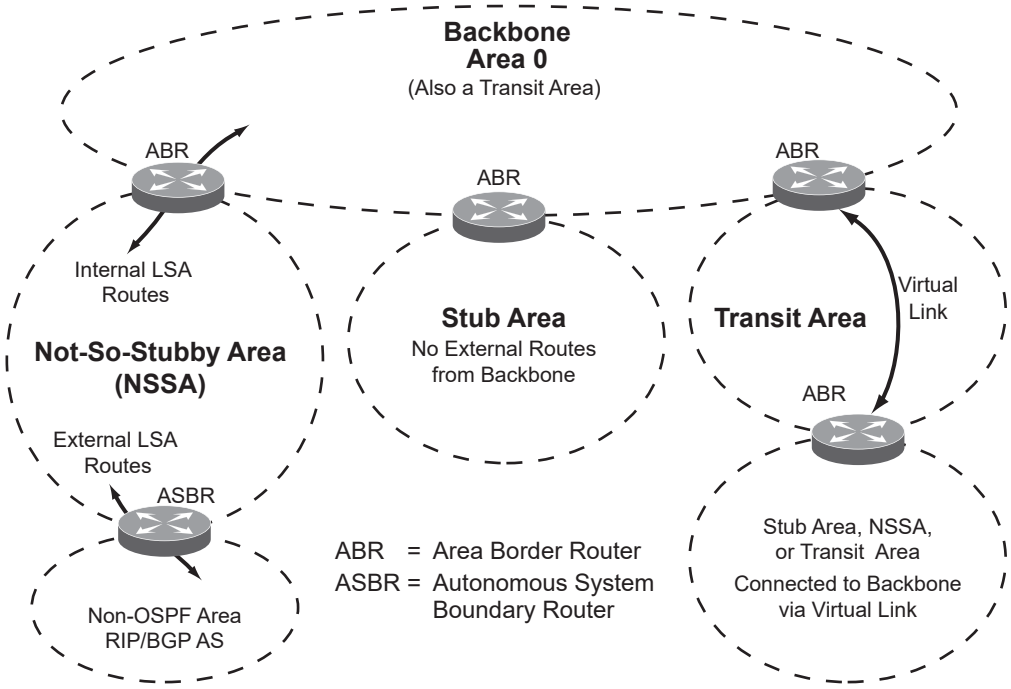
Types of OSPF Areas

An AS can be broken into logical units known as *areas*. In any AS with multiple areas, one area must be designated as area 0, known as the *backbone*. The backbone acts as the central OSPF area. All other areas in the AS must be connected to the backbone. Areas inject summary routing information into the backbone, which then distributes it to other areas as needed.

As shown in Figure 47, OSPF defines the following types of areas:

- Stub Area—an area that is connected to only one other area. External route information is not distributed into stub areas.
- Not-So-Stubby-Area (NSSA)—similar to a stub area with additional capabilities. Routes originating from within the NSSA can be propagated to adjacent transit and backbone areas. External routes from outside the AS can be advertised within the NSSA but can be configured to not be distributed into other areas.
- Transit Area—an area that carries data traffic which neither originates nor terminates in the area itself.

Figure 47. OSPF Area Types

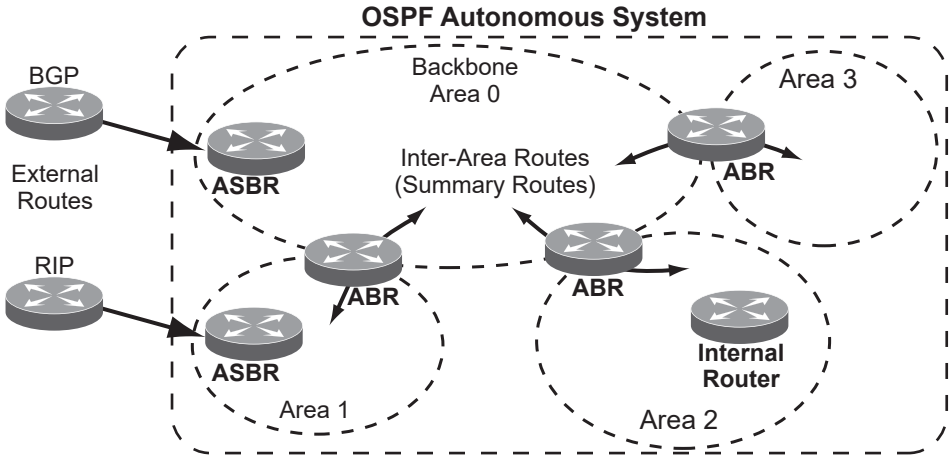


Types of OSPF Routing Devices

As shown in Figure 48, OSPF uses the following types of routing devices:

- Internal Router (IR)—a router that has all of its interfaces within the same area. IRs maintain LSDBs identical to those of other routing devices within the local area.
- Area Border Router (ABR)—a router that has interfaces in multiple areas. ABRs maintain one LSDB for each connected area and disseminate routing information between areas.
- Autonomous System Boundary Router (ASBR)—a router that acts as a gateway between the OSPF domain and non-OSPF domains, such as RIP, BGP, and static routes.

Figure 48. OSPF Domain and an Autonomous System



Neighbors and Adjacencies

In areas with two or more routing devices, *neighbors* and *adjacencies* are formed.

Neighbors are routing devices that maintain information about each others' health. To establish neighbor relationships, routing devices periodically send hello packets on each of their interfaces. All routing devices that share a common network segment, appear in the same area, and have the same health parameters (`hello` and `dead` intervals) and authentication parameters respond to each other's hello packets and become neighbors. Neighbors continue to send periodic hello packets to advertise their health to neighbors. In turn, they listen to hello packets to determine the health of their neighbors and to establish contact with new neighbors.

The hello process is used for electing one of the neighbors as the area's Designated Router (DR) and one as the area's Backup Designated Router (BDR). The DR is adjacent to all other neighbors and acts as the central contact for database exchanges. Each neighbor sends its database information to the DR, which relays the information to the other neighbors.

The BDR is adjacent to all other neighbors (including the DR). Each neighbor sends its database information to the BDR just as with the DR, but the BDR merely stores this data and does not distribute it. If the DR fails, the BDR will take over the task of distributing database information to the other neighbors.

The Link-State Database

OSPF is a link-state routing protocol. A *link* represents an interface (or routable path) from the routing device. By establishing an adjacency with the DR, each routing device in an OSPF area maintains an identical Link-State Database (LSDB) describing the network topology for its area.

Each routing device transmits a Link-State Advertisement (LSA) on each of its *active* interfaces. LSAs are entered into the LSDB of each routing device. OSPF uses *flooding* to distribute LSAs between routing devices. Interfaces may also be *passive*. Passive interfaces send LSAs to active interfaces, but do not receive LSAs, hello packets, or any other OSPF protocol information from active interfaces. Passive interfaces behave as stub networks, allowing OSPF routing devices to be aware of devices that do otherwise participate in OSPF (either because they do not support it, or because the administrator chooses to restrict OSPF traffic exchange or transit).

When LSAs result in changes to the routing device's LSDB, the routing device forwards the changes to the adjacent neighbors (the DR and BDR) for distribution to the other neighbors.

OSPF routing updates occur only when changes occur, instead of periodically. For each new route, if an adjacency is interested in that route (for example, if configured to receive static routes and the new route is indeed static), an update message containing the new route is sent to the adjacency. For each route removed from the route table, if the route has already been sent to an adjacency, an update message containing the route to withdraw is sent.

The Shortest Path First Tree

The routing devices use a link-state algorithm (Dijkstra's algorithm) to calculate the shortest path to all known destinations, based on the cumulative *cost* required to reach the destination.

The cost of an individual interface in OSPF is an indication of the overhead required to send packets across it. The cost is inversely proportional to the bandwidth of the interface. A lower cost indicates a higher bandwidth.

Internal Versus External Routing

To ensure effective processing of network traffic, every routing device on your network needs to know how to send a packet (directly or indirectly) to any other location/destination in your network. This is referred to as *internal routing* and can be done with static routes or using active internal routing protocols, such as OSPF, RIP, or RIPv2.

It is also useful to tell routers outside your network (upstream providers or *peers*) about the routes you have access to in your network. Sharing of routing information between autonomous systems is known as *external routing*.

Typically, an AS will have one or more border routers (peer routers that exchange routes with other OSPF networks) as well as an internal routing system enabling every router in that AS to reach every other router and destination within that AS.

When a routing device *advertises* routes to boundary routers on other autonomous systems, it is effectively committing to carry data to the IP space represented in the route being advertised. For example, if the routing device advertises 192.204.4.0/24, it is declaring that if another router sends data destined for any address in the 192.204.4.0/24 range, it will carry that data to its destination.

OSPFv2 Implementation in Enterprise NOS

Enterprise NOS supports a single instance of OSPF and up to 2K routes on the network. The following sections describe OSPF implementation in Enterprise NOS:

Configurable Parameters

In Enterprise NOS, OSPF parameters can be configured through the Command Line Interfaces (CLI/ISCLI), Browser-Based Interface (BBI), or through SNMP. For more information, see "Switch Administration" on page 29."

The CLI supports the following parameters: interface output cost, interface priority, dead and hello intervals, retransmission interval, and interface transmit delay.

In addition to the preceding parameters, you can specify the following:

- Shortest Path First (SPF) interval—Time interval between successive calculations of the shortest path tree using the Dijkstra's algorithm.
- Stub area metric—A stub area can be configured to send a numeric metric value such that all routes received via that stub area carry the configured metric to potentially influence routing decisions.
- Default routes—Default routes with weight metrics can be manually injected into transit areas. This helps establish a preferred route when multiple routing devices exist between two areas. It also helps route traffic to external networks.
- Passive—When enabled, the interface sends LSAs to upstream devices, but does not otherwise participate in OSPF protocol exchanges.
- Point-to-Point—For LANs that have only two OSPF routing agents (the CN4093 and one other device), this option allows the switch to significantly reduce the amount of routing information it must carry and manage.

Defining Areas

If you are configuring multiple areas in your OSPF domain, one of the areas must be designated as area 0, known as the *backbone*. The backbone is the central OSPF area and is usually physically connected to all other areas. The areas inject routing information into the backbone which, in turn, disseminates the information into other areas.

Since the backbone connects the areas in your network, it must be a contiguous area. If the backbone is partitioned (possibly as a result of joining separate OSPF networks), parts of the AS will be unreachable, and you will need to configure *virtual links* to reconnect the partitioned areas (see "Virtual Links" on page 481).

Up to three OSPF areas can be connected to the CN4093 with Enterprise NOS software. To configure an area, the OSPF number must be defined and then attached to a network interface on the switch. The full process is explained in the following sections.

An OSPF area is defined by assigning ***two*** pieces of information: an *area index* and an *area ID*. The commands to define and enable an OSPF area are as follows:

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area <area index> area-id <n.n.n.n>
CN 4093(config-router-ospf)# area <area index> enable
CN 4093(config-router-ospf)# exit
```

Note: The `aindex` option above is an arbitrary index used only on the switch and does not represent the actual OSPF area number. The actual OSPF area number is defined in the `areaid` portion of the command as explained in the following sections.

Assigning the Area Index

The `aindex` <*area index*> option is actually just an arbitrary index (0-2) used only by the CN4093. This index does not necessarily represent the OSPF area number, though for configuration simplicity, it should where possible.

For example, both of the following sets of commands define OSPF area 0 (the backbone) and area 1 because that information is held in the area ID portion of the command. However, the first set of commands is easier to maintain because the arbitrary area indexes agree with the area IDs:

- Area index and area ID agree

  `area 0 area-id 0.0.0.0` *(Use index 0 to set area 0 in ID octet format)*

  `area 1 area-id 0.0.0.1` *(Use index 1 to set area 1 in ID octet format)*

- Area index set to an arbitrary value

  `area 1 area-id 0.0.0.0` *(Use index 1 to set area 0 in ID octet format)*

  `area 2 area-id 0.0.0.1` *(Use index 2 to set area 1 in ID octet format)*

Using the Area ID to Assign the OSPF Area Number

The OSPF area number is defined in the `areaid` <*IP address*> option. The octet format is used to be compatible with two different systems of notation used by other OSPF network vendors. There are two valid ways to designate an area ID:

- Single Number

  Most common OSPF vendors express the area ID number as a single number. For example, the Cisco IOS-based router command "`network 1.1.1.0 0.0.0.255 area 1`" defines the area number simply as "`area 1.`"

- Multi-octet (*IP address*): Placing the area number in the last octet (0.0.0.*n*)

  Some OSPF vendors express the area ID number in multi-octet format. For example, "`area 0.0.0.2`" represents OSPF area 2 and can be specified directly on the CN4093 as "`area-id 0.0.0.2`".

On the CN4093, using the last octet in the area ID, "`area 1`" is equivalent to "`area-id 0.0.0.1`".

Note: Although both types of area ID formats are supported, be sure that the area IDs are in the same format throughout an area.

Attaching an Area to a Network

Once an OSPF area has been defined, it must be associated with a network. To attach the area to a network, you must assign the OSPF area index to an IP interface that participates in the area. The commands are as follows:

```
CN 4093(config)# interface ip <interface number>
CN 4093(config-ip-if)# ip ospf area <area index>
CN 4093(config-ip-if)# exit
```

For example, the following commands could be used to configure IPv4 interface 14 for a presence on the IPv4 10.10.10.1/24 network, to define OSPF area 1, and to attach the area to the network:

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area 1 area-id 0.0.0.1
CN 4093(config-router-ospf)# area 1 enable
CN 4093(config-router-ospf)# enable
CN 4093(config-router-ospf)# exit
CN 4093(config)# interface ip 14
CN 4093(config-ip-if)# ip address 10.10.10.1 255.255.255.0 enable
CN 4093(config-ip-if)# ip ospf area 1
CN 4093(config-ip-if)# ip ospf enable
```

Note: OSPFv2 supports IPv4 only. IPv6 is supported in OSPFv3 (see "OSPFv3 Implementation in Enterprise NOS" on page 495).

Interface Cost

The OSPF link-state algorithm (Dijkstra's algorithm) places each routing device at the root of a tree and determines the cumulative *cost* required to reach each destination. Usually, the cost is inversely proportional to the bandwidth of the interface. Low cost indicates high bandwidth. You can manually enter the cost for the output route with the following command:

```
CN 4093(config-ip-if)# ip ospf cost <cost value (1-65535)>
```

Electing the Designated Router and Backup

In any area with more than two routing devices, a Designated Router (DR) is elected as the central contact for database exchanges among neighbors, and a Backup Designated Router (BDR) is elected in case the DR fails.

DR and BDR elections are made through the hello process. The election can be influenced by assigning a priority value to the OSPF interfaces on the CN4093. The command is as follows:

```
CN 4093(config-ip-if)# ip ospf priority <priority value (0-255)>
```

A priority value of 255 is the highest, and 1 is the lowest. A priority value of 0 specifies that the interface cannot be used as a DR or BDR. In case of a tie, the routing device with the highest router ID wins. Interfaces configured as *passive* do not participate in the DR or BDR election process:

```
CN 4093(config-ip-if)# ip ospf passive-interface
CN 4093(config-ip-if)# exit
```

Summarizing Routes

Route summarization condenses routing information. Without summarization, each routing device in an OSPF network would retain a route to every subnet in the network. With summarization, routing devices can reduce some sets of routes to a single advertisement, reducing both the load on the routing device and the perceived complexity of the network. The importance of route summarization increases with network size.

Summary routes can be defined for up to 16 IP address ranges using the following command:

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area-range <range number> address <IP address>
<mask>
```

where <*range number*> is a number 1 to 16, <*IPv4 address*> is the base IP address for the range, and <*subnet mask*> is the IPv4 address mask for the range. For a detailed configuration example, see "Example 3: Summarizing Routes" on page 492.

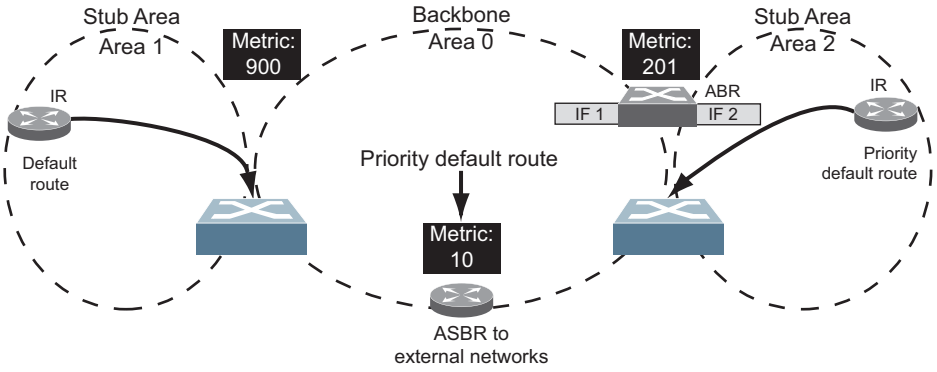
Note: OSPFv2 supports IPv4 only. IPv6 is supported in OSPFv3 (see "OSPFv3 Implementation in Enterprise NOS" on page 495).

Default Routes

When an OSPF routing device encounters traffic for a destination address it does not recognize, it forwards that traffic along the *default route*. Typically, the default route leads upstream toward the backbone until it reaches the intended area or an external router.

Each CN4093 acting as an ABR automatically inserts a default route into each attached area. In simple OSPF stub areas or NSSAs with only one ABR leading upstream (see Area 1 in Figure 49), any traffic for IP address destinations outside the area is forwarded to the switch's IP interface, and then into the connected transit area (usually the backbone). Since this is automatic, no further configuration is required for such areas.

Figure 49. Injecting Default Routes



If the switch is in a transit area and has a configured default gateway, it can inject a default route into rest of the OSPF domain. Use the following command to configure the switch to inject OSPF default routes:

```
CN 4093(config-router-ospf)# default-information <metric value> <metric type>
```

In the command above, *<metric value>* sets the priority for choosing this switch for default route. The value none sets no default and 1 sets the highest priority for default route. Metric type determines the method for influencing routing decisions for external routes.

When the switch is configured to inject a default route, an AS-external LSA with link state ID 0.0.0.0 is propagated throughout the OSPF routing domain. This LSA is sent with the configured metric value and metric type.

The OSPF default route configuration can be removed with the command:

```
CN 4093(config-router-ospf)# no default-information
```

Virtual Links

Usually, all areas in an OSPF AS are physically connected to the backbone. In some cases where this is not possible, you can use a *virtual link*. Virtual links are created to connect one area to the backbone through another non-backbone area (see Figure 47 on page 472).

The area which contains a virtual link must be a transit area and have full routing information. Virtual links cannot be configured inside a stub area or NSSA. The area type must be defined as `transit` using the following command:

```
CN 4093(config-router-ospf)# area <area index> type transit
```

The virtual link must be configured on the routing devices at each endpoint of the virtual link, though they may traverse multiple routing devices. To configure a CN4093 as one endpoint of a virtual link, use the following command:

```
CN 4093(config-router-ospf)# area-virtual-link <link number> neighbor-router
<router ID>
```

where <*link number*> is a value between 1 and 3, <*area index*> is the OSPF area index of the transit area, and <*router ID*> is the IP address of the virtual neighbor (nbr), the routing device at the target endpoint. Another router ID is needed when configuring a virtual link in the other direction. To provide the CN4093 with a router ID, see the following section, Router ID.

For a detailed configuration example on Virtual Links, see "Example 2: Virtual Links" on page 489.

Router ID

Routing devices in OSPF areas are identified by a router ID, expressed in IP address format. The router ID is not required to be part of any IP interface range or in any OSPF area, and may even use the CN4093 loopback interface (see "Loopback Interfaces in OSPF" on page 485).

The router ID can be configured in one of the following two ways:

- Dynamically (the default)—OSPF protocol configures the router ID as the lowest IP loopback interface IP address, if available, or else the lowest IP interface IP address, if available. Once dynamically configured, the router ID does not normally undergo further updates.
- Statically—Use the following command to manually configure the router ID:

  ```
  CN 4093(config-router-ospf)# ip router-id <IPv4 address>
  ```

To change the router ID from static to dynamic, set the router ID to 0.0.0.0, save the configuration, and reboot the CN4093. To view the router ID, enter:

```
CN 4093(config-router-ospf)# show ip ospf
```

Authentication

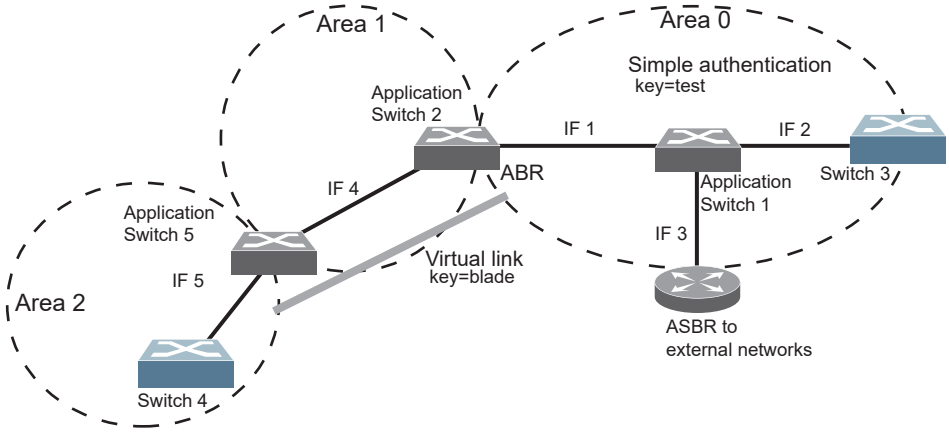
OSPF protocol exchanges can be authenticated so that only trusted routing devices can participate. This ensures less processing on routing devices that are not listening to OSPF packets.

OSPF allows packet authentication and uses IP multicast when sending and receiving packets. Routers participate in routing domains based on pre-defined passwords. Enterprise NOS supports simple password (type 1 plain text passwords) and MD5 cryptographic authentication. This type of authentication allows a password to be configured per area.

We strongly recommend that you implement MD5 cryptographic authentication as a best practice.

Figure shows authentication configured for area 0 with the password `test`. Simple authentication is also configured for the virtual link between area 2 and area 0. Area 1 is not configured for OSPF authentication.

Figure 50. OSPF Authentication



Configuring Plain Text OSPF Passwords

To configure plain text OSPF passwords as shown in Figure use the following commands:

1. Enable OSPF authentication for Area 0 on switches 1, 2, and 3.

```
CN 4093(config-router-ospf)# area 0 authentication-type password
CN 4093(config-router-ospf)# exit
```

2. Configure a simple text password up to eight characters for each OSPF IP interface in Area 0 on switches 1, 2, and 3.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip ospf key test
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip ospf key test
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ip ospf key test
CN 4093(config-ip-if)# exit
```

3. Enable OSPF authentication for Area 2 on switch 4.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area 2 authentication-type password
```

4. Configure a simple text password up to eight characters for the virtual link between Area 2 and Area 0 on switches 2 and 4.

```
CN 4093(config-router-ospf)# area-virtual-link 1 key IBM
```

Configuring MD5 Authentication

Use the following commands to configure MD5 authentication on the switches shown in Figure :

1. Enable OSPF MD5 authentication for Area 0 on switches 1, 2, and 3.

```
CN 4093(config-router-ospf)# area 0 authentication-type md5
```

2. Configure MD5 key ID for Area 0 on switches 1, 2, and 3.

```
CN 4093(config-router-ospf)# message-digest-key 1 md5-key test
CN 4093(config-router-ospf)# exit
```

3. Assign MD5 key ID to OSPF interfaces on switches 1, 2, and 3.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip ospf message-digest-key 1
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip ospf message-digest-key 1
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ip ospf message-digest-key 1
CN 4093(config-ip-if)# exit
```

4. Enable OSPF MD5 authentication for Area 2 on switch 4.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area 2 authentication-type md5
```

5. Configure MD5 key for the virtual link between Area 2 and Area 0 on switch 2 and switch 4.

```
CN 4093(config-router-ospf)# message-digest-key 2 md5-key test
```

6. Assign MD5 key ID to OSPF virtual link on switches 2 and 4.

```
CN 4093(config-router-ospf)# area-virtual-link 1 message-digest-key 2
CN 4093(config-router-ospf)# exit
```

Host Routes for Load Balancing

Enterprise NOS implementation of OSPF includes host routes. Host routes are used for advertising network device IP addresses to external networks, accomplishing the following goals:

- ABR Load Sharing

  As a form of load balancing, host routes can be used for dividing OSPF traffic among multiple ABRs. To accomplish this, each switch provides identical services but advertises a host route for a different IP address to the external network. If each IP address serves a different and equal portion of the external world, incoming traffic from the upstream router should be split evenly among ABRs.

- ABR Failover

  Complementing ABR load sharing, identical host routes can be configured on each ABR. These host routes can be given different costs so that a different ABR is selected as the preferred route for each server and the others are available as backups for failover purposes.

- Equal Cost Multipath (ECMP)

  With equal cost multipath, a router potentially has several available next hops towards any given destination. ECMP allows separate routes to be calculated for each IP Type of Service. All paths of equal cost to a given destination are calculated, and the next hops for all equal-cost paths are inserted into the routing table.

If redundant routes via multiple routing processes (such as OSPF, RIP, BGP, or static routes) exist on your network, the switch defaults to the OSPF-derived route.

Loopback Interfaces in OSPF

Because loopback interfaces are always available on the switch, loopback interfaces may present an advantage when used as the router ID.

If dynamic router ID selection is used (see "Router ID" on page 481), loopback interfaces can be used to force router ID selection. If a loopback interface is configured, its IP address is automatically selected as the router ID, even if other IP interfaces have lower IP addresses. If more than one loopback interface is configured, the lowest loopback interface IP address is selected.

Loopback interfaces can be advertised into the OSPF domain by specifying an OSPF host route with the loopback interface IP address.

To enable OSPF on an existing loopback interface:

```
CN 4093(config)# interface loopback <1-5>
CN 4093(config-ip-loopback)# ip ospf area <area ID> enable
CN 4093(config-ip-loopback)# exit
```

OSPF Features Not Supported

The following OSPF features are not supported in this release:

- Summarizing external routes
- Filtering OSPF routes
- Using OSPF to forward multicast routes
- Configuring OSPF on non-broadcast multi-access networks (such as frame relay, X.25, or ATM)

OSPFv2 Configuration Examples

A summary of the basic steps for configuring OSPF on the CN4093 is listed here. Detailed instructions for each of the steps is covered in the following sections:

1. Configure IP interfaces.

   One IP interface is required for each desired network (range of IP addresses) being assigned to an OSPF area on the switch.

2. (Optional) Configure the router ID.

   The router ID is required only when configuring virtual links on the switch.

3. Enable OSPF on the switch.
4. Define the OSPF areas.
5. Configure OSPF interface parameters.

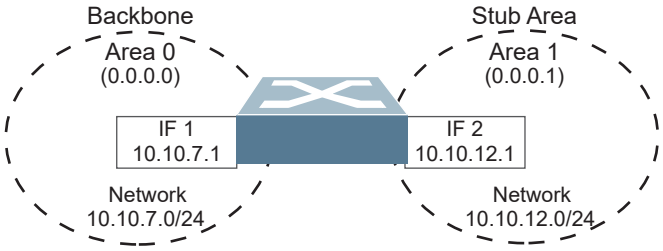
   IP interfaces are used for attaching networks to the various areas.

6. (Optional) Configure route summarization between OSPF areas.
7. (Optional) Configure virtual links.
8. (Optional) Configure host routes.

Example 1: Simple OSPF Domain

In this example, two OSPF areas are defined—one area is the backbone and the other is a stub area. A stub area does not allow advertisements of external routes, thus reducing the size of the database. Instead, a default summary route of IP address 0.0.0.0 is automatically inserted into the stub area. Any traffic for IP address destinations outside the stub area will be forwarded to the stub area's IP interface, and then into the backbone.

Figure 51. A Simple OSPF Domain



Follow this procedure to configure OSPF support as shown in Figure 51:

1. Configure IP interfaces on each network that will be attached to OSPF areas.

   In this example, two IP interfaces are needed:

   - Interface 1 for the backbone network on 10.10.7.0/24
   - Interface 2 for the stub area network on 10.10.12.0/24

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 10.10.7.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 10.10.12.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
```

   Note: OSPFv2 supports IPv4 only. IPv6 is supported in OSPFv3 (see "OSPFv3 Implementation in Enterprise NOS" on page 495).

2. Enable OSPF.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# enable
```

3. Define the backbone.

   The backbone is always configured as a transit area using `areaid 0.0.0.0`.

```
CN 4093(config-router-ospf)# area 0 area-id 0.0.0.0
CN 4093(config-router-ospf)# area 0 type transit
CN 4093(config-router-ospf)# area 0 enable
```

4. Define the stub area.

```
CN 4093(config-router-ospf)# area 1 area-id 0.0.0.1
CN 4093(config-router-ospf)# area 1 type stub
CN 4093(config-router-ospf)# area 1 enable
CN 4093(config-router-ospf)# exit
```

5. Attach the network interface to the backbone.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip ospf area 0
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

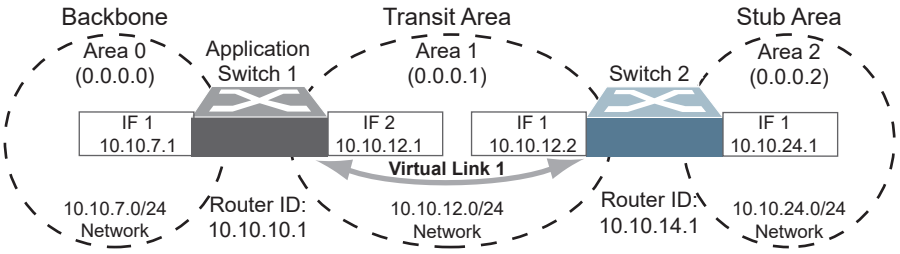
6. Attach the network interface to the stub area.

```
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip ospf area 1
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

Example 2: Virtual Links

In the example shown in Figure 52, area 2 is not physically connected to the backbone as is usually required. Instead, area 2 will be connected to the backbone via a virtual link through area 1. The virtual link must be configured at each endpoint.

Figure 52. Configuring a Virtual Link



Note: OSPFv2 supports IPv4 only. IPv6 is supported in OSPFv3 (see "OSPFv3 Implementation in Enterprise NOS" on page 495).

Configuring OSPF for a Virtual Link on Switch #1

1. Configure IP interfaces on each network that will be attached to the switch.

   In this example, two IP interfaces are needed:

   - Interface 1 for the backbone network on 10.10.7.0/24
   - Interface 2 for the transit area network on 10.10.12.0/24

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 10.10.7.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 10.10.12.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
```

2. Configure the router ID.

   A router ID is required when configuring virtual links. Later, when configuring the other end of the virtual link on Switch 2, the router ID specified here will be used as the target virtual neighbor (`nbr`) address.

```
CN 4093(config)# ip router-id 10.10.10.1
```

3. Enable OSPF.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# enable
```

4. Define the backbone.

```
CN 4093(config-router-ospf)# area 0 area-id 0.0.0.0
CN 4093(config-router-ospf)# area 0 type transit
CN 4093(config-router-ospf)# area 0 enable
```

5. Define the transit area.

   The area that contains the virtual link must be configured as a transit area.

```
CN 4093(config-router-ospf)# area 1 area-id 0.0.0.1
CN 4093(config-router-ospf)# area 1 type transit
CN 4093(config-router-ospf)# area 1 enable
CN 4093(config-router-ospf)# exit
```

6. Attach the network interface to the backbone.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip ospf area 0
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

7. Attach the network interface to the transit area.

```
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip ospf area 1
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

8. Configure the virtual link.

   The `nbr` router ID configured in this step must be the same as the router ID that will be configured for Switch #2 in Step 2 on page 491.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area-virtual-link 1 area 1
CN 4093(config-router-ospf)# area-virtual-link 1 neighbor-router
      10.10.14.1
CN 4093(config-router-ospf)# area-virtual-link 1 enable
```

Configuring OSPF for a Virtual Link on Switch #2

1. Configure IP interfaces on each network that will be attached to OSPF areas.

   In this example, two IP interfaces are needed:

   - Interface 1 for the transit area network on 10.10.12.0/24
   - Interface 2 for the stub area network on 10.10.24.0/24

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 10.10.12.2 255.255.255.0 enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 10.10.24.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
```

2. Configure the router ID.

   A router ID is required when configuring virtual links. This router ID should be the same one specified as the target virtual neighbor (`nbr`) on switch 1 in Step 8 on page 490.

```
CN 4093(config)# ip router-id 10.10.14.1
```

3. Enable OSPF.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# enable
```

4. Define the backbone.

   This version of Enterprise NOS requires that a backbone index be configured on the non-backbone end of the virtual link as follows:

```
CN 4093(config-router-ospf)# area 0 area-id 0.0.0.0
CN 4093(config-router-ospf)# area 0 enable
```

5. Define the transit area.

```
CN 4093(config-router-ospf)# area 1 area-id 0.0.0.1
CN 4093(config-router-ospf)# area 1 type transit
CN 4093(config-router-ospf)# area 1 enable
```

6. Define the stub area.

```
CN 4093(config-router-ospf)# area 2 area-id 0.0.0.2
CN 4093(config-router-ospf)# area 2 type stub
CN 4093(config-router-ospf)# area 2 enable
CN 4093(config-router-ospf)# exit
```

7. Attach the network interface to the transmit area:

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip ospf area 1
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

8. Attach the network interface to the stub area.

```
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip ospf area 2
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

9. Configure the virtual link.

   The `nbr` router ID configured in this step must be the same as the router ID that was configured for switch #1 in Step 2 on page 489.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area-virtual-link 1 area 1
CN 4093(config-router-ospf)# area-virtual-link 1 neighbor-router
      10.10.10.1
CN 4093(config-router-ospf)# area-virtual-link 1 enable
```

Other Virtual Link Options

- You can use redundant paths by configuring multiple virtual links.
- Only the endpoints of the virtual link are configured. The virtual link path may traverse multiple routers in an area as long as there is a routable path between the endpoints.

Example 3: Summarizing Routes

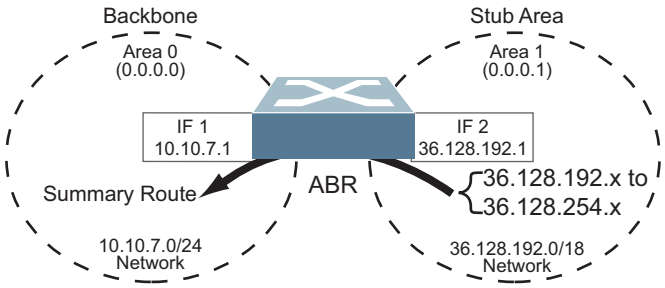
By default, ABRs advertise all the network addresses from one area into another area. Route summarization can be used for consolidating advertised addresses and reducing the perceived complexity of the network.

If the network IP addresses in an area are assigned to a contiguous subnet range, you can configure the ABR to advertise a single summary route that includes all the individual IP addresses within the area.

The following example shows one summary route from area 1 (stub area) injected into area 0 (the backbone). The summary route consists of all IP addresses from 36.128.192.0 through 36.128.254.255 except for the routes in the range 36.128.200.0 through 36.128.200.255.

Note: OSPFv2 supports IPv4 only. IPv6 is supported in OSPFv3 (see "OSPFv3 Implementation in Enterprise NOS" on page 495).

Figure 53. Summarizing Routes



Note: You can specify a range of addresses to prevent advertising by using the hide option. In this example, routes in the range 36.128.200.0 through 36.128.200.255 are kept private.

Follow this procedure to configure OSPF support as shown in Figure 53:

1. Configure IP interfaces for each network which will be attached to OSPF areas.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 10.10.7.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 36.128.192.1 255.255.255.0 enable
CN 4093(config-ip-if)# exit
```

2. Enable OSPF.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# enable
```

3. Define the backbone.

```
CN 4093(config-router-ospf)# area 0 area-id 0.0.0.0
CN 4093(config-router-ospf)# area 0 type transit
CN 4093(config-router-ospf)# area 0 enable
```

4. Define the stub area.

```
CN 4093(config-router-ospf)# area 1 area-id 0.0.0.1
CN 4093(config-router-ospf)# area 1 type stub
CN 4093(config-router-ospf)# area 1 enable
CN 4093(config-router-ospf)# exit
```

5. Attach the network interface to the backbone.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip ospf area 0
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

6. Attach the network interface to the stub area.

```
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip ospf area 1
CN 4093(config-ip-if)# ip ospf enable
CN 4093(config-ip-if)# exit
```

7. Configure route summarization by specifying the starting address and mask of the range of addresses to be summarized.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area-range 1 address 36.128.192.0
      255.255.192.0
CN 4093(config-router-ospf)# area-range 1 area 1
CN 4093(config-router-ospf)# area-range 1 enable
CN 4093(config-router-ospf)# exit
```

8. Use the hide command to prevent a range of addresses from advertising to the backbone.

```
CN 4093(config)# router ospf
CN 4093(config-router-ospf)# area-range 2 address 36.128.200.0
255.255.255.0
CN 4093(config-router-ospf)# area-range 2 area 1
CN 4093(config-router-ospf)# area-range 2 hide
CN 4093(config-router-ospf)# exit
```

Verifying OSPF Configuration

Use the following commands to verify the OSPF configuration on your switch:

- **show ip ospf**
- **show ip ospf neighbor**
- **show ip ospf database database-summary**
- **show ip ospf routes**

Refer to the *Enterprise NOS Command Reference* for information on the preceding commands.

OSPFv3 Implementation in Enterprise NOS

OSPF version 3 is based on OSPF version 2, but has been modified to support IPv6 addressing. In most other ways, OSPFv3 is similar to OSPFv2: They both have the same packet types and interfaces, and both use the same mechanisms for neighbor discovery, adjacency formation, LSA flooding, aging, and so on. The administrator should be familiar with the OSPFv2 concepts covered in the preceding sections of this chapter before implementing the OSPFv3 differences as described in the following sections.

Although OSPFv2 and OSPFv3 are very similar, they represent independent features on the CN4093. They are configured separately, and both can run in parallel on the switch with no relation to one another, serving different IPv6 and IPv4 traffic, respectively.

OSPFv3 Differences from OSPFv2

Note: When OSPFv3 is enabled, the OSPF backbone area (0.0.0.0) is created by default and is always active.

OSPFv3 Requires IPv6 Interfaces

OSPFv3 is designed to support IPv6 addresses. This requires IPv6 interfaces to be configured on the switch and assigned to OSPF areas, in much the same way IPv4 interfaces are assigned to areas in OSPFv2. This is the primary configuration difference between OSPFv3 and OSPFv2.

See "Internet Protocol Version 6" on page 407 for configuring IPv6 interfaces.

OSPFv3 Uses Independent Command Paths

Though OSPFv3 and OSPFv2 are very similar, they are configured independently. OSPFv3 command paths are located as follows:

- In the ISCLI

```
CN 4093(config)# ipv6 router ospf                    (OSPFv3 router config mode)
CN 4093(config-router-ospf3)# ?

CN 4093(config)# interface ip <Interface number>     (Configure OSPFv3)
CN 4093(config-ip-if)# ipv6 ospf ?                   (OSPFv3 interface config)

CN 4093# show ipv6 ospf ?                             (Show OSPFv3 information)
```

OSPFv3 Identifies Neighbors by Router ID

Where OSPFv2 uses a mix of IPv4 interface addresses and Router IDs to identify neighbors, depending on their type, OSPFv3 configuration consistently uses a Router ID to identify all neighbors.

Although Router IDs are written in dotted decimal notation, and may even be based on IPv4 addresses from an original OSPFv2 network configuration, it is important to realize that Router IDs are not IP addresses in OSPFv3, and can be assigned independently of IP address space. However, maintaining Router IDs consistent with any legacy OSPFv2 IPv4 addressing allows for easier implementation of both protocols.

Other Internal Improvements

OSPFv3 has numerous improvements that increase the protocol efficiency in addition to supporting IPv6 addressing. These improvements change some of the behaviors in the OSPFv3 network and may affect topology consideration, but have little direct impact on configuration. For example:

- Addressing fields have been removed from Router and Network LSAs.
- Flexible treatment of unknown LSA types to make integration of OSPFv3 easier.
- Interface network type can be specified using the command:
  `CN 4093(config-ip-if)# ipv6 ospf network {broadcast| |non-broadcast|point-to-multipoint|point-to-point}`
- For an interface network type that is not broadcast or NBMA, link LSA suppression can be enabled so link LSA is not originated for the interface. Use the command: `CN 4093(config-ip-if)# ipv6 ospf linklsasuppress`

OSPFv3 Limitations

Enterprise NOS 8.4 does not currently support the following OSPFv3 features:

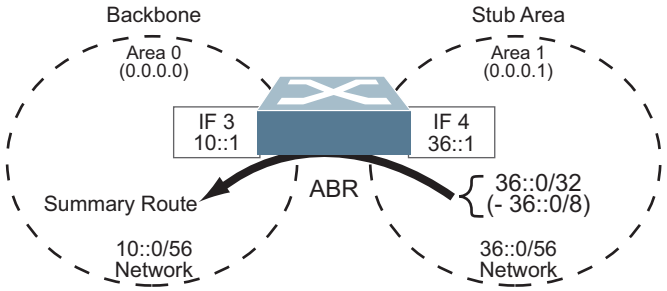
- Multiple instances of OSPFv3 on one IPv6 link.

OSPFv3 Configuration Example

The following example depicts the OSPFv3 equivalent configuration of "Example 3: Summarizing Routes" on page 492 for OSPFv2.

In this example, one summary route from area 1 (stub area) is injected into area 0 (the backbone). The summary route consists of all IP addresses for the 36.:0/32 portion of the 36::0/56 network except for the routes in the 36::0/8 range.

Figure 54. Summarizing Routes



Note: You can specify a range of addresses to prevent advertising by using the hide option. In this example, routes in the 36::0/8 range are kept private.

Use the following procedure to configure OSPFv3 support as shown in Figure 53:

1. Configure IPv6 interfaces for each link which will be attached to OSPFv3 areas.

```
CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ipv6 address 10:0:0:0:0:0:0:1
CN 4093(config-ip-if)# ipv6 prefixlen 56
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 4
CN 4093(config-ip-if)# ip address 36:0:0:0:0:0:1
CN 4093(config-ip-if)# ipv6 prefixlen 56
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
```

   This is equivalent to configuring the IP address and netmask for IPv4 interfaces.

2. Enable OSPFv3.

```
CN 4093(config)# ipv6 router ospf
CN 4093(config-router-ospf3)# enable
```

   This is equivalent to the OSPFv2 `enable` option in the `router ospf` command path.

3. Define the backbone.

```
CN 4093(config-router-ospf3)# area 0 area-id 0.0.0.0
CN 4093(config-router-ospf3)# area 0 type transit
CN 4093(config-router-ospf3)# area 0 enable
```

   This is identical to OSPFv2 configuration.

4. Define the stub area.

```
CN 4093(config-router-ospf3)# area 1 area-id 0.0.0.1
CN 4093(config-router-ospf3)# area 1 type stub
CN 4093(config-router-ospf3)# area 1 enable
CN 4093(config-router-ospf3)# exit
```

   This is identical to OSPFv2 configuration.

5. Attach the network interface to the backbone.

```
CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ipv6 ospf area 0
CN 4093(config-ip-if)# ipv6 ospf enable
CN 4093(config-ip-if)# exit
```

   The `ipv6` command path is used instead of the OSPFv2 `ip` command path

6. Attach the network interface to the stub area.

```
CN 4093(config)# interface ip 4
CN 4093(config-ip-if)# ipv6 ospf area 1
CN 4093(config-ip-if)# ipv6 ospf enable
CN 4093(config-ip-if)# exit
```

   The `ipv6` command path is used instead of the OSPFv2 `ip` command path.

7. Configure route summarization by specifying the starting address and prefix length of the range of addresses to be summarized.

```
CN 4093(config)# ipv6 router ospf
CN 4093(config-router-ospf3)# area-range 1 address 36:0:0:0:0:0:0:0 32
CN 4093(config-router-ospf3)# area-range 1 area 0
CN 4093(config-router-ospf3)# area-range 1 enable
```

   This differs from OSPFv2 only in that the OSPFv3 command path is used, and the address and prefix are specified in IPv6 format.

8. Use the hide command to prevent a range of addresses from advertising to the backbone.

```
CN 4093(config-router-ospf)# area-range 2 address 36:0:0:0:0:0:0:0 8
CN 4093(config-router-ospf)# area-range 2 area 0
CN 4093(config-router-ospf)# area-range 2 hide
CN 4093(config-router-ospf)# exit
```

   This differs from OSPFv2 only in that the OSPFv3 command path is used, and the address and prefix are specified in IPv6 format.

Neighbor Configuration Example

When using NBMA or point to multipoint interfaces, you must manually configure neighbors. The following example includes the steps for neighbor configuration.

1. Configure IPv6 interface parameters:

```
CN 4093(config# interface ip 10
CN 4093(config-ip-if)# ipv6 address 10:0:0:0:0:0:0:12 64
CN 4093(config-ip-if)# vlan 10
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# ipv6 ospf area 0
CN 4093(config-ip-if)# ipv6 ospf retransmit-interval 5
CN 4093(config-ip-if)# ipv6 ospf transmit-delay 1
CN 4093(config-ip-if)# ipv6 ospf priority 1
CN 4093(config-ip-if)# ipv6 ospf hello-interval 10
CN 4093(config-ip-if)# ipv6 ospf dead-interval 40
CN 4093(config-ip-if)# ipv6 ospf network point-to-multipoint
CN 4093(config-ip-if)# ipv6 ospf poll-interval 120
CN 4093(config-ip-if)# ipv6 ospf enable
CN 4093(config-ip-if)# exit
```

2. Enable OSPFv3:

```
CN 4093(config# ipv6 router ospf
CN 4093(config-router-ospf3)# router-id 12.12.12.12
CN 4093(config-router-ospf3)# enable
```

3. Define the backbone.

```
CN 4093(config-router-ospf3)# area 0 area-id 0.0.0.0
CN 4093(config-router-ospf3)# area 0 stability-interval 40
CN 4093(config-router-ospf3)# area 0 default-metric 1
CN 4093(config-router-ospf3)# area 0 default-metric type 1
CN 4093(config-router-ospf3)# area 0 translation-role candidate
CN 4093(config-router-ospf3)# area 0 type transit
CN 4093(config-router-ospf3)# area 0 enable
```

4. Configure neighbor entry:

```
CN 4093(config-router-ospf3)# neighbor 1 address
fe80:0:0:0:dceb:ff:fe00:9
CN 4093(config-router-ospf3)# neighbor 1 interface 10
CN 4093(config-router-ospf3)# neighbor 1 priority 1
CN 4093(config-router-ospf3)# neighbor 1 enable
```

Chapter 31. Protocol Independent Multicast

Enterprise NOS supports Protocol Independent Multicast (PIM) in Sparse Mode (PIM-SM) and Dense Mode (PIM-DM).

Note: Enterprise NOS 8.4 does not support IPv6 for PIM.

The following sections discuss PIM support for the CN4093 10 Gb Converged Scalable Switch:

PIM Overview

PIM is designed for efficiently routing multicast traffic across one or more IPv4 domains. This has benefits for application such as IP television, collaboration, education, and software delivery, where a single source must deliver content (a multicast) to a group of receivers that span both wide-area and inter-domain networks.

Instead of sending a separate copy of content to each receiver, a multicast derives efficiency by sending only a single copy of content toward its intended receivers. This single copy only becomes duplicated when it reaches the target domain that includes multiple receivers, or when it reaches a necessary bifurcation point leading to different receiver domains.

PIM is used by multicast source stations, client receivers, and intermediary routers and switches, to build and maintain efficient multicast routing trees. PIM is protocol independent; It collects routing information using the existing unicast routing functions underlying the IPv4 network, but does not rely on any particular unicast protocol. For PIM to function, a Layer 3 routing protocol (such as BGP, OSPF, RIP, or static routes) must first be configured on the switch.

PIM-SM is a reverse-path routing mechanism. Client receiver stations advertise their willingness to join a multicast group. The local routing and switching devices collect multicast routing information and forward the request toward the station that will provide the multicast content. When the join requests reach the sending station, the multicast data is sent toward the receivers, flowing in the opposite direction of the original join requests.

Some routing and switching devices perform special PIM-SM functions. Within each receiver domain, one router is elected as the Designated Router (DR) for handling multicasts for the domain. DRs forward information to a similar device, the Rendezvous Point (RP), which holds the root tree for the particular multicast group.

Receiver join requests as well as sender multicast content initially converge at the RP, which generates and distributes multicast routing data for the DRs along the delivery path. As the multicast content flows, DRs use the routing tree information obtained from the RP to optimize the paths both to and from send and receive stations, bypassing the RP for the remainder of content transactions if a more efficient route is available.

DRs continue to share routing information with the RP, modifying the multicast routing tree when new receivers join, or pruning the tree when all the receivers in any particular domain are no longer part of the multicast group.

Supported PIM Modes and Features

For each interface attached to a PIM network component, PIM can be configured to operate either in PIM Sparse Mode (PIM-SM) or PIM Dense Mode (PIM-DM).

- PIM-SM is used in networks where multicast senders and receivers comprise a relatively small (sparse) portion of the overall network. PIM-SM uses a more complex process than PIM-DM for collecting and optimizing multicast routes, but minimizes impact on other IP services and is more commonly used.
- PIM-DM is used where multicast devices are a relatively large (dense) portion of the network, with very frequent (or constant) multicast traffic. PIM-DM requires less configuration on the switch than PIM-SM, but uses broadcasts that can consume more bandwidth in establishing and optimizing routes.

The following PIM modes and features are *not* currently supported in Enterprise NOS 8.4:

- Hybrid Sparse-Dense Mode (PIM-SM/DM). Sparse Mode and Dense Mode may be configured on separate IP interfaces on the switch, but are not currently supported simultaneously on the same IP interface.
- PIM Source-Specific Multicast (PIM-SSM)
- Anycast RP
- PIM RP filters
- Only configuration via the switch ISCLI is supported. PIM configuration is currently not available using the menu-based CLI, the BBI, or via SNMP.

Basic PIM Settings

To use PIM the following is required:

- The PIM feature must be enabled globally on the switch.
- PIM network components and PIM modes must be defined.
- IP interfaces must be configured for each PIM component.
- PIM neighbor filters may be defined (optional).
- If PIM-SM is used, define additional parameters:
  - Rendezvous Point
  - Designated Router preferences (optional)
  - Bootstrap Router preferences (optional)

Each of these tasks is covered in the following sections.

Note: PIM can be configured through the ISCLI only. PIM configuration and information are not available using the menu-based CLI, the BBI, or via SNMP.

Globally Enabling or Disabling the PIM Feature

By default, PIM is disabled on the switch. PIM can be globally enabled or disabled using the following ISCLI commands:

```
CN 4093(config)# [no] ip pim enable
```

Defining a PIM Network Component

The CN4093 can be attached to a maximum of two independent PIM network components. Each component represents a different PIM network, and can be defined for either PIM-SM or PIM-DM operation. Basic PIM component configuration is performed using the following commands:

```
CN 4093(config)# ip pim component <1-2>
CN 4093(config-ip-pim-comp)# mode {sparse|dense}
CN 4093(config-ip-pim-comp)# exit
```

The `sparse` option will place the component in Sparse Mode (PIM-SM). The `dense` option will place the component in Dense Mode (PIM-DM). By default, PIM component 1 is configured for Sparse Mode. PIM component 2 is non configured by default.

Note: A component using PIM-SM must also be configured with a dynamic or static Rendezvous Point (see "Specifying the Rendezvous Point" on page 506).

Defining an IP Interface for PIM Use

Each network attached to an IP interface on the switch may be assigned one of the available PIM components. The same PIM component can be assigned to multiple IP interfaces. The interfaces may belong to the same VLAN, but each interface can belong to only one VLAN.

To define an IP interface for use with PIM, first configured the interface with an IPv4 address and VLAN as follows:

```
CN 4093(config)# interface ip <Interface number>
CN 4093(config-ip-if)# ip address <IPv4 address> <IPv4 mask>
CN 4093(config-ip-if)# vlan <VLAN number>
CN 4093(config-ip-if)# enable
```

Note: The PIM feature currently supports only one VLAN for each IP interface. Configurations where different interfaces on different VLANs share IP addresses are not supported.

Next, PIM must be enabled on the interface, and the PIM network component ID must be specified:

```
CN 4093(config-ip-if)# ip pim enable
CN 4093(config-ip-if)# ip pim component-id <1-2>
CN 4093(config-ip-if)# exit
```

By default, PIM component 1 is automatically assigned when PIM is enabled on the IP interface.

Note: While PIM is enabled on the interface, the interface VLAN cannot be changed. To change the VLAN, first disable PIM on the interface.

PIM Neighbor Filters

The CN4093 accepts connection to up to 24 PIM interfaces. By default, the switch accepts all PIM neighbors attached to the PIM-enabled interfaces, up to the maximum number (72 neighbors). Once the maximum is reached, the switch will deny further PIM neighbors.

To ensure that only the appropriate PIM neighbors are accepted by the switch, the administrator can use PIM neighbor filters to specify which PIM neighbors may be accepted or denied on a per-interface basis.

To turn PIM neighbor filtering on or off for a particular IP interface, use the following commands:

```
CN 4093(config)# interface ip <Interface number>
CN 4093(config-ip-if)# [no] ip pim neighbor-filter
```

When filtering is enabled, all PIM neighbor requests on the specified IP interface will be denied by default. To allow a specific PIM neighbor, use the following command:

```
CN 4093(config-ip-if)# ip pim neighbor-addr <neighbor IPv4 address> allow
```

To remove a PIM neighbor from the accepted list, use the following command.

```
CN 4093(config-ip-if)# ip pim neighbor-addr <neighbor IPv4 address> deny
CN 4093(config-ip-if)# exit
```

You can view configured PIM neighbor filters globally or for a specific IP interface using the following commands:

```
CN 4093(config)# show ip pim neighbor-filters
CN 4093(config)# show ip pim interface <Interface number> neighbor-filters
```

Additional Sparse Mode Settings

Specifying the Rendezvous Point

Using PIM-SM, at least one PIM-capable router must be a candidate for use as a Rendezvous Point (RP) for any given multicast group. If desired, the CN4093 can act as an RP candidate. To assign a configured switch IP interface as a candidate, use the following procedure.

1. Select the PIM component that will represent the RP candidate:

```
CN 4093(config)# ip pim component <1-2>
```

2. Configure the IPv4 address of the switch interface which will be advertised as a candidate RP for the specified multicast group:

```
CN 4093(config-ip-pim-comp)# rp-candidate rp-address <group address>
<group address mask> <candidate IPv4 address>
```

The switch interface will participate in the election of the RP that occurs on the Bootstrap Router, or BSR (see "Specifying a Bootstrap Router" on page 507).

Alternately, if no election is desired, the switch can provide a static RP, specified using the following command:

```
CN 4093(config-ip-pim-comp)# rp-static rp-address <group address>
<group address mask> <static RP IPv4 address>
```

3. If using dynamic RP candidates, configure the amount of time that the elected interface will remain the RP for the group before a re-election is performed:

```
CN 4093(config-ip-pim-comp)# rp-candidate holdtime <0-255>
CN 4093(config-ip-pim-comp)# exit
```

Influencing the Designated Router Selection

Using PIM-SM, All PIM-enabled IP interfaces are considered as potential Designate Routers (DR) for their domain. By default, the interface with the highest IP address on the domain is selected. However, if an interface is configured with a DR priority value, it overrides the IP address selection process. If more than one interface on a domain is configured with a DR priority, the one with the highest number is selected.

Use the following commands to configure the DR priority value (Interface IP mode):

```
CN 4093(config)# interface ip <Interface number>
CN 4093(config-ip-if)# ip pim dr-priority <value (0-4294967294)>
CN 4093(config-ip-if)# exit
```

Note: A value of 0 (zero) specifies that the CN4093 will not act as the DR. This setting requires the CN4093 to be connected to a peer that has a DR priority setting of 1 or higher in order to ensure that a DR will be present in the network.

Specifying a Bootstrap Router

Using PIM-SM, a Bootstrap Router (BSR) is a PIM-capable router that hosts the election of the RP from available candidate routers. For each PIM-enabled IP interface, the administrator can set the preference level for which the local interface becomes the BSR:

```
CN 4093(config)# interface ip <Interface number>
CN 4093(config-ip-if)# ip pim cbsr-preference <0 to 255>
CN 4093(config-ip-if)# exit
```

A value of 255 highly prefers the local interface as a BSR. A value of -1 indicates that the PIM CBSR preference is not configured on the local interface.

Using PIM with Other Features

PIM with ACLs or VMAPs

If using ACLs or VMAPs, be sure to permit traffic for local hosts and routers.

PIM with IGMP

If using IGMP (see "Internet Group Management Protocol" on page 439):

- IGMP static joins can be configured with a PIM-SM or PIM-DM multicast group IPv4 address. Using the ISCLI:

```
CN 4093(config)# ip mroute <multicast group IPv4 address> <VLAN> <port>
```

- IGMP Query is disabled by default. If IGMP Querier is needed with PIM, be sure to enable the IGMP Query feature globally, as well as on each VLAN where it is needed.
- If the switch is connected to multicast receivers and/or hosts, be sure to enable IGMP snooping globally, as well as on each VLAN where PIM receivers are attached.

PIM Configuration Examples

Example 1: PIM-SM with Dynamic RP

This example configures PIM Sparse Mode for one IP interface, with the switch acting as a candidate for dynamic Rendezvous Point (RP) selection.

1. Globally enable the PIM feature:

```
CN 4093(config)# ip pim enable
```

2. Configure a PIM network component with dynamic RP settings, and set it for PIM Sparse Mode:

```
CN 4093(config)# ip pim component 1
CN 4093(config-ip-pim-comp)# mode sparse
CN 4093(config-ip-pim-comp)# rp-candidate rp-address 225.1.0.0
255.255.0.0 10.10.1.1
CN 4093(config-ip-pim-comp)# exit
```

Where 225.1.0.0 is the multicast group base IP address, 255.255.0.0 is the multicast group address mask, and 10.10.1.1 is the switch RP candidate address.

Note: Because, Sparse Mode is set by default for PIM component 1, the `mode` command is needed only if the mode has been previously changed.

3. Define an IP interface for use with PIM:

```
CN 4093(config)# interface ip 111
CN 4093(config-ip-if)# ip address 10.10.1.1 255.255.255.255
CN 4093(config-ip-if)# vlan 11
CN 4093(config-ip-if)# enable
```

The IP interface represents the PIM network being connected to the switch. The IPv4 addresses in the defined range must not be included in another IP interface on the switch under a different VLAN.

4. Enable PIM on the IP interface and assign the PIM component:

```
CN 4093(config-ip-if)# ip pim enable
CN 4093(config-ip-if)# ip pim component-id 1
```

Note: Because, PIM component 1 is assigned to the interface by default, the `component-id` command is needed only if the setting has been previously changed.

5. Set the Bootstrap Router (BSR) preference:

```
CN 4093(config-ip-if)# ip pim cbsr-preference 135
CN 4093(config-ip-if)# exit
```

Example 2: PIM-SM with Static RP

The following commands can be used to modify the prior example configuration to use a static RP:

```
CN 4093(config)# ip pim static-rp enable
CN 4093(config)# ip pim component 1
CN 4093(config-ip-pim-comp)# rp-static rp-address 225.1.0.0 255.255.0.0
10.10.1.1
CN 4093(config-ip-pim-comp)# exit
```

Where 225.1.0.0 255.255.0.0 is the multicast group base address and mask, and 10.10.1.1 is the static RP address.

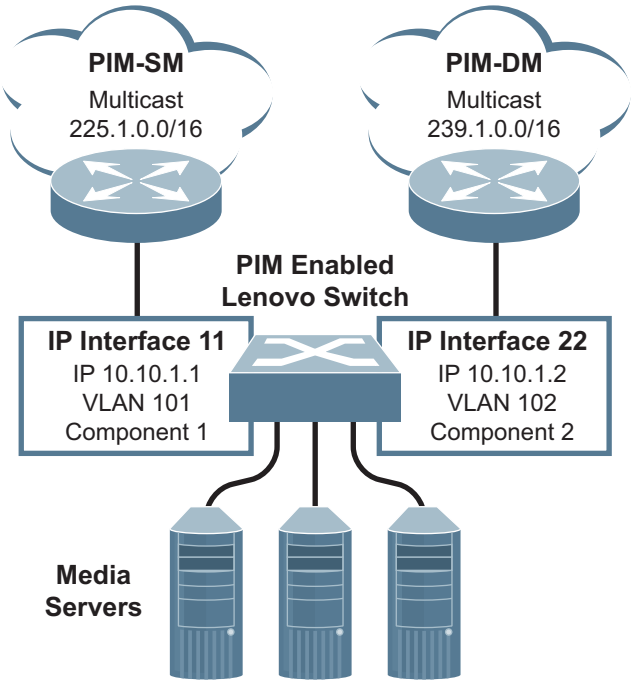
Note: The same static RP address should be configured for all switches in the group.

Example 3: PIM-DM

This example configures PIM Dense Mode (PIM-DM) on one IP interface. PIM-DM can be configured independently, or it can be combined with the prior PIM-SM examples (which are configured on a different PIM component) as shown in Figure 55.

Note: In the following example, since the receivers and sources are connected in different areas, the border router must be configured for the IPMC traffic to be forwarded. Enterprise NOS supports only partial configuration of PIM border router.

Figure 55. Network with both PIM-DM and PIM-SM Components



1. Configure the PIM-SM component as shown in the prior examples, or if using PIM-DM independently, enable the PIM feature.

```
CN 4093(config)# ip pim enable
```

2. Configure a PIM component and set the PIM mode:

```
CN 4093(config)# ip pim component 2
CN 4093(config-ip-pim-comp)# mode dense
CN 4093(config-ip-pim-comp)# exit
```

3. Define an IP interface for use with PIM:

```
CN 4093(config)# interface ip 22
CN 4093(config-ip-if)# ip address 10.10.1.2 255.255.255.255
CN 4093(config-ip-if)# vlan 102
CN 4093(config-ip-if)# enable
```

4. Enable PIM on the IP interface and assign the PIM component:

```
CN 4093(config-ip-if)# ip pim enable
CN 4093(config-ip-if)# ip pim component-id 2
CN 4093(config-ip-if)# exit
```

5. (Optional) Configure PIM border router if the IPMC traffic is flowing between PIM domains:

```
CN 4093(config)# ip pim pmbr enable
CN 4093(config)# interface ip 22
CN 4093(config-ip-if)# ip pim border-bit
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 11
CN 4093(config-ip-if)# ip pim border-bit
CN 4093(config-ip-if)# exit
```

Note: For PIM Dense Mode, the DR, RP, and BSR settings do not apply.

Part 6: High Availability Fundamentals

Internet traffic consists of myriad services and applications which use the Internet Protocol (IP) for data delivery. However, IP is not optimized for all the various applications. High Availability goes beyond IP and makes intelligent switching decisions to provide redundant network configurations.

Chapter 32. Basic Redundancy

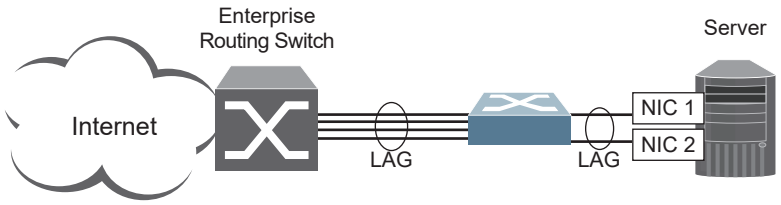
Enterprise NOS 8.4 includes various features for providing basic link or device redundancy:

Aggregation for Link Redundancy

Multiple switch ports can be combined together to form robust, high-bandwidth LAGs to other devices. Since LAGs are comprised of multiple physical links, the LAG is inherently fault tolerant. As long as one connection between the switches is available, the LAG remains active.

In Figure 56, four ports are aggregated together between the switch and the enterprise routing device. Connectivity is maintained as long as one of the links remains active. The links to the server are also aggregated, allowing the secondary NIC to take over in the event that the primary NIC link fails.

Figure 56. Aggregating Ports for Link Redundancy



For more information about aggregation, see "Ports and Link Aggregation (LAG)" on page 161.

Hot Links

Hot Links provides basic link redundancy with fast recovery.

Hot Links consists of up to 200 triggers. A trigger consists of a pair of layer 2 interfaces, each containing an individual port, LAG, or LACP adminkey. One interface is the Master, and the other is a Backup. While the Master interface is set to the active state and forwards traffic, the Backup interface is set to the standby state and blocks traffic until the Master interface fails. If the Master interface fails, the Backup interface is set to active and forwards traffic. Once the Master interface is restored, it transitions to the standby state and blocks traffic until the Backup interface fails.

You may select a physical port, static LAG, or an LACP adminkey as a Hot Link interface. Only external uplink ports can be members of a Hot Links trigger interface.

Forward Delay

The Forward Delay timer allows Hot Links to monitor the Master and Backup interfaces for link stability before selecting one interface to transition to the active state. Before the transition occurs, the interface must maintain a stable link for the duration of the Forward Delay interval.

For example, if you set the Forward delay timer to 10 seconds using the command:

```
CN 4093(config)# hotlinks trigger <x> forward-delay 10
```

the switch will select an interface to become active only if a link remained stable for the duration of the Forward Delay period. If the link is unstable, the Forward Delay period starts again.

Preemption

You can configure the Master interface to resume the active state whenever it becomes available. With Hot Links preemption enabled, the Master interface transitions to the active state immediately upon recovery. The Backup interface immediately transitions to the standby state. If Forward Delay is enabled, the transition occurs when an interface has maintained link stability for the duration of the Forward Delay period.

FDB Update

Use the FDB update option to notify other devices on the network about updates to the Forwarding Database (FDB). When you enable FDB update, the switch sends multicasts of addresses in the forwarding database (FDB) over the active interface, so that other devices on the network can learn the new path. The Hot Links FBD update option uses the station update rate to determine the rate at which to send FDB packets.

Configuration Guidelines

The following configuration guidelines apply to Hot links:

- Only external ports can be configured as Hot Links.
- When Hot Links is turned on, MSTP, RSTP, and PVRST must be turned off.
- A port that is a member of the Master interface cannot be a member of the Backup interface. A port that is a member of one Hot Links trigger cannot be a member of another Hot Links trigger.
- An individual port that is configured as a Hot Link interface cannot be a member of a LAG.

Configuring Hot Links

Use the following commands to configure Hot Links.

```
CN 4093(config)# hotlinks trigger 1 enable   (Enable Hot Links Trigger 1)
CN 4093(config)# hotlinks trigger 1 master port 38  (Add port to Master interface)
CN 4093(config)# hotlinks trigger 1 backup port 39  (Add port to Backup interface)
CN 4093(config)# hotlinks enable       (Turn on Hot Links)
```

Chapter 33. Layer 2 Failover

The primary application for Layer 2 Failover is to support Network Adapter Teaming. With Network Adapter Teaming, all the NICs on each server share the same IP address, and are configured into a team. One NIC is the primary link, and the other is a standby link.For more details, refer to the documentation for your Ethernet adapter.

Note: Only two links per server blade can be used for Layer 2 LAG Failover (one primary and one backup). Network Adapter Teaming allows only one backup NIC for each server blade.

Auto Monitoring LAG Links

Layer 2 Failover can be enabled on any LAG in the CN4093, including LACP LAGs. LAGs can be added to failover trigger groups. Then, if some specified number of trigger links fail, the switch disables all the internal ports in the switch (unless VLAN Monitor is turned on). When the internal ports are disabled, it causes the NIC team on the affected server blades to failover from the primary to the backup NIC. This process is called a failover event.

When the appropriate number of links in a trigger group return to service, the switch enables the internal ports. This causes the NIC team on the affected server blades to fail back to the primary switch (unless Auto-Fallback is disabled on the NIC team). The backup switch processes traffic until the primary switch's internal links come up, which can take up to five seconds.

VLAN Monitor

The VLAN Monitor allows Layer 2 Failover to discern different VLANs. With VLAN Monitor turned on:

- If enough links in a trigger fail (see "Setting the Failover Limit" on page 522), the switch disables all internal ports that reside in the same VLAN membership as the LAG(s) in the trigger.
- When enough links in the trigger return to service, the switch enables the internal ports that reside in the same VLAN membership as the LAG(s) in the trigger.

If you turn off the VLAN Monitor (`CN 4093# no failover vlan`), only one failover trigger is allowed. When a link failure occurs on the trigger, the switch disables all internal server-blade ports.

Auto Monitor Configurations

Figure 57 is an example of Layer 2 Failover. One CN4093 is the primary and the other is used as a backup. In this example, all external ports on the primary switch belong to a single LAG, with Layer 2 Failover enabled and Failover Limit set to 2. If two or fewer links in trigger 1 remain active, the switch temporarily disables all internal server-blade ports that reside in VLAN 1. This action causes a failover event on Server 1 and Server 2.

Figure 57. Basic Layer 2 Failover

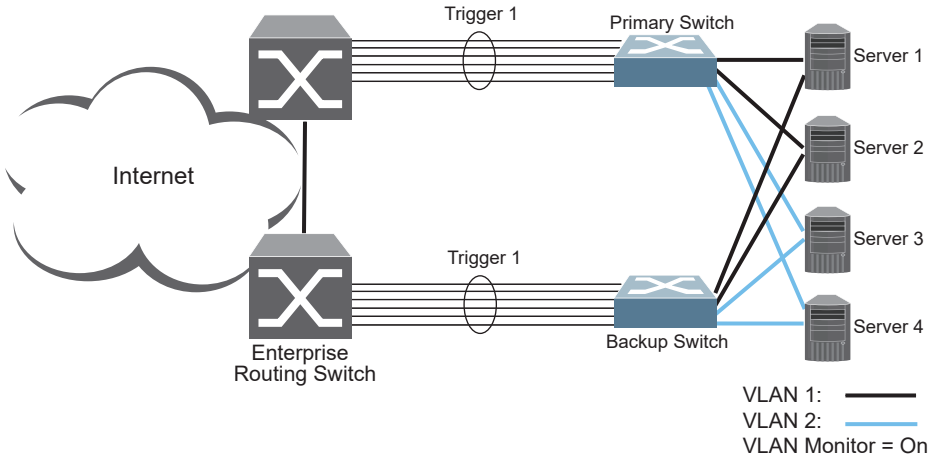


Figure 58 shows a configuration with two LAGs, each in a different Failover Trigger. Switch 1 is the primary switch for Server 1 and Server 2. Switch 2 is the primary switch for Server 3 and Server 4. VLAN Monitor is turned on. STP is turned off.

If all links go down in trigger 1, Switch 1 disables all internal ports that reside in VLAN 1. If all links in trigger 2 go down, Switch 1 disables all internal ports that reside in VLAN 2.

Figure 58. Two LAGs, each in a different Failover Trigger

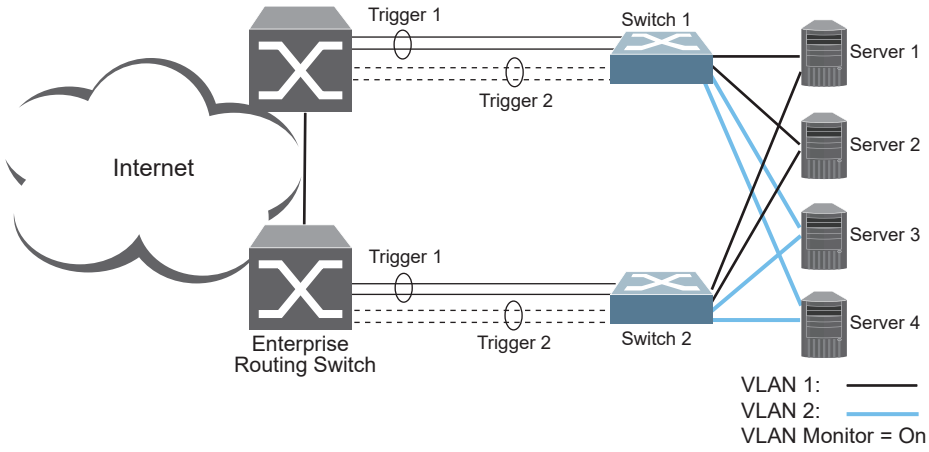
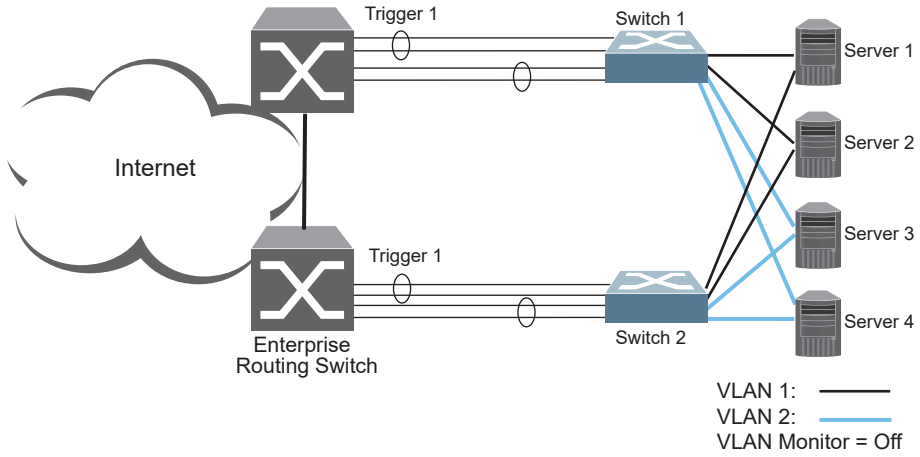


Figure 59 shows a configuration with two LAGs. VLAN Monitor is turned off, so only one Failover Trigger is configured on each switch. Switch 1 is the primary switch for Server 1 and Server 2. Switch 2 is the primary switch for Server 3 and Server 4. STP is turned off.

If all links in trigger 1 go down, switch 1 disables all internal links to server blades.

Figure 59. Two LAGs, one Failover Trigger



Setting the Failover Limit

The failover limit lets you specify the minimum number of operational links required within each trigger before the trigger initiates a failover event. For example, if the limit is two, a failover event occurs when the number of operational links in the trigger is two or fewer. When you set the limit to zero, the switch triggers a failover event only when no links in the trigger are operational.

Manually Monitoring Port Links

The Manual Monitor allows you to configure a set of ports and/or LAGs to monitor for link failures (a monitor list) and another set of ports and/or LAGs to disable when the trigger limit is reached (a control list). When the switch detects a link failure on the monitor list, it automatically disables the items in control list. When server ports are disabled, the corresponding server's network adapter can detect the disabled link and trigger a network-adapter failover to another port or LAG on the switch or another switch in the chassis.

The switch automatically enables the control list items when the monitor list items return to service.

Monitor Port State

A monitor port is considered operational as long as the following conditions are true:

- The port must be in the `Link Up` state.
- If STP is enabled, the port must be in the `Forwarding` state.
- If the port is part of an LACP LAG, the port must be in the `Aggregated` state.

If any of the above conditions is false, the monitor port is considered to have failed.

Control Port State

A control port is considered Operational if the monitor trigger is up. As long as the trigger is up, the port is considered operational from a teaming perspective, even if the port itself is actually in the `Down` state, `Blocking` state (if STP is enabled on the port), or `Not Aggregated` state (if part of an LACP LAG).

A control port is considered to have failed only if the monitor trigger is in the `Down` state.

To view the state of any port, use one of the following commands:

```
CN 4093# show interface link              (View port link status)
CN 4093# show interface port <x> spanning-tree stp <x>    (View port STP status)
CN 4093# show lacp information            (View port LACP status)
```

L2 Failover with Other Features

L2 Failover works together with Link Aggregation Control Protocol (LACP) and with Spanning Tree Protocol (STP), as described in the next sections.

LACP

Link Aggregation Control Protocol allows the switch to form dynamic LAGs. You can use the *admin key* to add up to two LACP LAGs to a failover trigger using automatic monitoring. When you add an *admin key* to a trigger, any LACP LAG with that *admin key* becomes a member of the trigger.

Note: If you change the LACP system priority on an LACP aggregation, the failover trigger goes down.

Spanning Tree Protocol

If Spanning Tree Protocol (STP) is enabled on the ports in a failover trigger, the switch monitors the port STP state rather than the link state. A port failure results when STP is not in a `Forwarding` state (such as `Learning`, `Discarding` or `No Link`). The switch automatically disables the appropriate internal ports, based on the VLAN monitor.

When the switch determines that ports in the trigger are in STP `Forwarding` state, then it automatically enables the appropriate internal ports, based on the VLAN monitor. The switch *fails back* to normal operation.

Configuration Guidelines

This section provides important information about configuring Layer 2 Failover.

Note: Auto Monitor and Manual Monitor are mutually exclusive. They cannot both be configured on the switch.

Auto Monitor Guidelines

- Any specific failover trigger may monitor static LAGs only or LACP LAGs only, but not both.
- All external ports in all static or LACP LAGs added to any specific failover trigger must belong to the same VLAN.
- A maximum of two LACP keys can be added per trigger.
- When VLAN Monitor is on, the following additional guidelines apply:
  - All external ports in all static or LACP LAGs added to a specific failover trigger must belong to the same VLAN and have the same PVID.
  - Different triggers are not permitted to operate on the same VLAN.
  - Different triggers are not permitted to operate on the same internal port.
  - For each port in each LAG in a specific failover trigger, the trigger will monitor the STP state on only the default PVID.

Manual Monitor Guidelines

- A Manual Monitor can monitor only external ports.
- Any specific failover trigger can monitor external ports only, static LAGs only, or LACP LAGs only. The different types cannot be combined in the same trigger.
- A maximum of two LACP keys can be added per trigger.
- Management ports, FC ports and stacking ports cannot be monitored.
- Control ports for different triggers must not overlap. Monitor ports may overlap.

Configuring Layer 2 Failover

Auto Monitor Example

The following procedure pertains to the configuration shown in Figure 57.

1. Configure Network Adapter Teaming on the servers.
2. Define a LAG on the CN4093.

```
CN 4093(config)# portchannel 1 port EXT1,EXT2,EXT3 enable
```

3. Configure Failover parameters.

```
CN 4093(config)# failover trigger 1 enable
CN 4093(config)# failover trigger 1 limit <0-1024>
CN 4093(config)# failover trigger 1 amon portchannel 1
```

4. Verify the configuration.

```
CN 4093(config)# show failover trigger 1 information
```

Manual Monitor Example

Use the following procedure to configure a Layer 2 Failover Manual Monitor.

1. Configure Network Adapter Teaming on the servers.
2. Specify the links to monitor.

```
CN 4093(config)# failover trigger 1 mmon monitor member EXT4,EXT5,EXT6
```

3. Specify the links to disable when the failover limit is reached.

```
CN 4093(config)# failover trigger 1 mmon control member INT13,INT14
```

4. Configure general Layer 2 Failover parameters.

```
CN 4093(config)# failover trigger 1 enable
CN 4093(config)# failover trigger 1 limit <0-1024>
```

5. Enable failover globally.

```
CN 4093(config)# failover enable
```

6. Verify the configuration.

```
CN 4093(config)# show failover trigger 1 information
```

Chapter 34. Virtual Router Redundancy Protocol

The CN4093 10 Gb Converged Scalable Switch (CN4093) supports IPv4 high-availability network topologies through an enhanced implementation of the Virtual Router Redundancy Protocol (VRRP).

Note: Enterprise NOS 8.4 does not support IPv6 for VRRP.

The following topics are discussed in this chapter:

VRRP Overview

In a high-availability network topology, no device can create a single point-of-failure for the network or force a single point-of-failure to any other part of the network. This means that your network will remain in service despite the failure of any single device. To achieve this usually requires redundancy for all vital network components.

VRRP enables redundant router configurations within a LAN, providing alternate router paths for a host to eliminate single points-of-failure within a network. Each participating VRRP-capable routing device is configured with the same virtual router IPv4 address and ID number. One of the virtual routers is elected as the master, based on a number of priority criteria, and assumes control of the shared virtual router IPv4 address. If the master fails, one of the backup virtual routers will take control of the virtual router IPv4 address and actively process traffic addressed to it.

With VRRP, Virtual Interface Routers (VIR) allow two VRRP routers to share an IP interface across the routers. VIRs provide a single Destination IPv4 (DIP) address for upstream routers to reach various servers, and provide a virtual default Gateway for the server blades.

VRRP Components

Each physical router running VRRP is known as a *VRRP router*.

Virtual Router

Two or more VRRP routers can be configured to form a *virtual router* (RFC 2338). Each VRRP router may participate in one or more virtual routers. Each virtual router consists of a user-configured *virtual router identifier* (VRID) and an IPv4 address.

Virtual Router MAC Address

The VRID is used to build the *virtual router MAC Address*. The five highest-order octets of the virtual router MAC Address are the standard MAC prefix (00-00-5E-00-01) defined in RFC 2338. The VRID is used to form the lowest-order octet.

Owners and Renters

Only one of the VRRP routers in a virtual router may be configured as the IPv4 address owner. This router has the virtual router's IPv4 address as its real interface address. This router responds to packets addressed to the virtual router's IPv4 address for ICMP pings, TCP connections, and so on.

There is no requirement for any VRRP router to be the IPv4 address owner. Most VRRP installations choose not to implement an IPv4 address owner. For the purposes of this chapter, VRRP routers that are not the IPv4 address owner are called *renters*.

Master and Backup Virtual Router

Within each virtual router, one VRRP router is selected to be the virtual router master. See "Selecting the Master VRRP Router" on page 529 for an explanation of the selection process.

Note: If the IPv4 address owner is available, it will always become the virtual router master.

The virtual router master forwards packets sent to the virtual router. It also responds to Address Resolution Protocol (ARP) requests sent to the virtual router's IPv4 address. Finally, the virtual router master sends out periodic advertisements to let other VRRP routers know it is alive and its priority.

Within a virtual router, the VRRP routers not selected to be the master are known as virtual router backups. Should the virtual router master fail, one of the virtual router backups becomes the master and assumes its responsibilities.

Virtual Interface Router

At Layer 3, a Virtual Interface Router (VIR) allows two VRRP routers to share an IP interface across the routers. VIRs provide a single Destination IPv4 (DIP) address for upstream routers to reach various destination networks, and provide a virtual default Gateway.

Note: Every VIR must be assigned to an IP interface, and every IP interface must be assigned to a VLAN. If no port in a VLAN has link up, the IP interface of that VLAN is down, and if the IP interface of a VIR is down, that VIR goes into INIT state.

VRRP Operation

Only the virtual router master responds to ARP requests. Therefore, the upstream routers only forward packets destined to the master. The master also responds to ICMP ping requests. The backup does not forward any traffic, nor does it respond to ARP requests.

If the master is not available, the backup becomes the master and takes over responsibility for packet forwarding and responding to ARP requests.

Selecting the Master VRRP Router

Each VRRP router is configured with a priority between 1–254. A bidding process determines which VRRP router is or becomes the master—the VRRP router with the highest priority.

The master periodically sends advertisements to an IPv4 multicast address. As long as the backups receive these advertisements, they remain in the backup state. If a backup does not receive an advertisement for three advertisement intervals, it initiates a bidding process to determine which VRRP router has the highest priority and takes over as master. In addition to the three advertisement intervals, a manually set holdoff time can further delay the backups from assuming the master status.

If, at any time, a backup determines that it has higher priority than the current master does, it can preempt the master and become the master itself, unless configured not to do so. In preemption, the backup assumes the role of master and begins to send its own advertisements. The current master sees that the backup has higher priority and will stop functioning as the master.

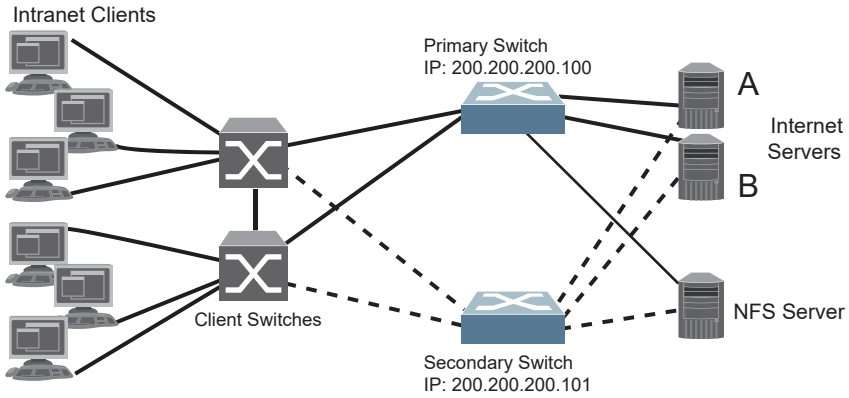
A backup router can stop receiving advertisements for one of two reasons—the master can be down, or all communications links between the master and the backup can be down. If the master has failed, it is clearly desirable for the backup (or one of the backups, if there is more than one) to become the master.

Note: If the master is healthy but communication between the master and the backup has failed, there will then be two masters within the virtual router. To prevent this from happening, configure redundant links to be used between the switches that form a virtual router.

Failover Methods

With service availability becoming a major concern on the Internet, service providers are increasingly deploying Internet traffic control devices, such as application switches, in redundant configurations. Traditionally, these configurations have been *hot-standby* configurations, where one switch is active and the other is in a standby mode. A non-VRRP hot-standby configuration is shown in the figure below:

Figure 60. A Non-VRRP, Hot-Standby Configuration



While hot-standby configurations increase site availability by removing single points-of-failure, service providers increasingly view them as an inefficient use of network resources because one functional application switch sits by idly until a failure calls it into action. Service providers now demand that vendors' equipment support redundant configurations where all devices can process traffic when they are healthy, increasing site throughput and decreasing user response times when no device has failed.

Enterprise NOS high availability configurations are based on VRRP. The implementation of VRRP includes proprietary extensions.

The Enterprise NOS implementation of VRRP supports the following modes of high availability:

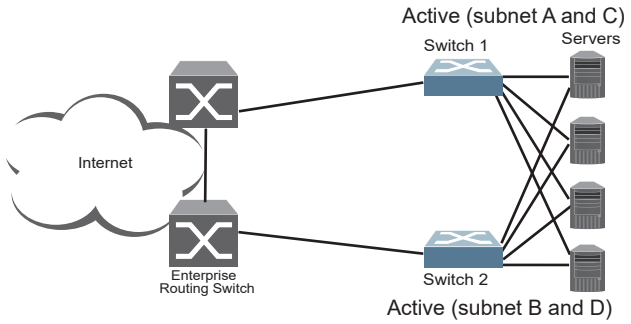
- **Active-Active**—based on proprietary Enterprise NOS extensions to VRRP
- **Hot-Standby**—supports Network Adapter Teaming on your server blades

Active-Active Redundancy

In an active-active configuration, shown in Figure 61, two switches provide redundancy for each other, with both active at the same time. Each switch processes traffic on a different subnet. When a failure occurs, the remaining switch can process traffic on all subnets.

For a configuration example, see "High Availability Configurations" on page 535.

Figure 61. Active-Active Redundancy

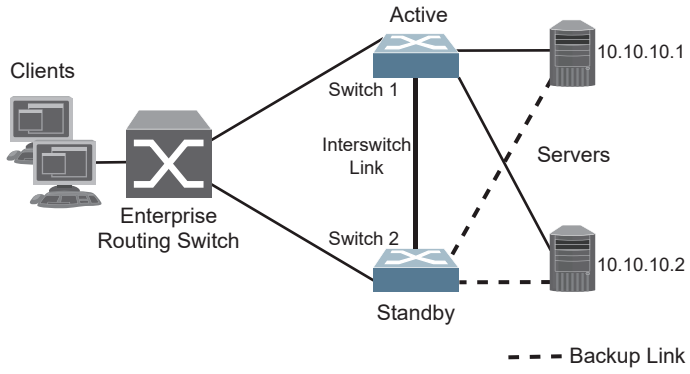


Hot-Standby Redundancy

The primary application for VRRP-based hot-standby is to support Server Load Balancing when you have configured Network Adapter Teaming on your server blades. With Network Adapter Teaming, the NICs on each server share the same IPv4 address, and are configured into a team. One NIC is the primary link, and the others are backup links. For more details, refer to the relevant network adapter documentation.

The hot-standby model is shown in Figure 62.

Figure 62. Hot-Standby Redundancy



Virtual Router Group

The virtual router group ties all virtual routers on the switch together as a single entity. By definition, hot-standby requires that all virtual routers failover as a group, and not individually. As members of a group, all virtual routers on the switch (and therefore the switch itself), are in either a master or standby state.

The virtual router group cannot be used for active-active configurations or any other configuration that require shared interfaces.

A VRRP group has the following characteristics:

- When enabled, all virtual routers behave as one entity, and all group settings override any individual virtual router settings.
- All individual virtual routers, once the VRRP group is enabled, assume the group's tracking and priority.
- When one member of a VRRP group fails, the priority of the group decreases, and the state of the entire switch changes from Master to Standby.

Each VRRP advertisement can include up to 128 addresses. All virtual routers are advertised within the same packet, conserving processing and buffering resources.

Enterprise NOS Extensions to VRRP

This section describes VRRP enhancements that are implemented in Enterprise NOS.

Enterprise NOS supports a tracking function that dynamically modifies the priority of a VRRP router, based on its current state. The objective of tracking is to have, whenever possible, the master bidding processes for various virtual routers in a LAN converge on the same switch. Tracking ensures that the selected switch is the one that offers optimal network performance. For tracking to have any effect on virtual router operation, preemption must be enabled.

Enterprise NOS can track the attributes listed in Table 38 :

Table 38. *VRRP Tracking Parameters*

Parameter	Description
Number of IP interfaces on the switch that are active (“up”) `tracking-priority-increment interfaces`	Helps elect the virtual routers with the most available routes as the master. (An IP interface is considered active when there is at least one active port on the same VLAN.) This parameter influences the VRRP router's priority in virtual interface routers.
Number of active ports on the same VLAN `tracking-priority-increment ports`	Helps elect the virtual routers with the most available ports as the master. This parameter influences the VRRP router's priority in virtual interface routers. **Note**: In a hot-standby configuration, only external ports are tracked.
Number of virtual routers in master mode on the switch `tracking-priority-increment virtual-routers`	Useful for ensuring that traffic for any particular client/server pair is handled by the same switch, increasing routing efficiency. This parameter influences the VRRP router's priority in virtual interface routers.

Each tracked parameter has a user-configurable weight associated with it. As the count associated with each tracked item increases (or decreases), so does the VRRP router's priority, subject to the weighting associated with each tracked item. If the priority level of a standby is greater than that of the current master, then the standby can assume the role of the master.

See “Configuring the Switch for Tracking” on page 534 for an example on how to configure the switch for tracking VRRP priority.

Virtual Router Deployment Considerations

Assigning VRRP Virtual Router ID

During the software upgrade process, VRRP virtual router IDs will be automatically assigned if failover is enabled on the switch. When configuring virtual routers at any point after upgrade, virtual router ID numbers must be assigned. The virtual router ID may be configured as any number between 1 and 255. Use the following commands to configure the virtual router ID:

```
CN 4093(config)# router vrrp
CN 4093(config-vrrp)# virtual-router 1 virtual-router-id <1-255>
```

Configuring the Switch for Tracking

Tracking configuration largely depends on user preferences and network environment. Consider the configuration shown in Figure 61 on page 531. Assume the following behavior on the network:

- Switch 1 is the master router upon initialization.
- If switch 1 is the master and it has one fewer active servers than switch 2, then switch 1 remains the master.

  This behavior is preferred because running one server down is less disruptive than bringing a new master online and severing all active connections in the process.
- If switch 1 is the master and it has two or more active servers fewer than switch 2, then switch 2 becomes the master.
- If switch 2 is the master, it remains the master even if servers are restored on switch 1 such that it has one fewer or an equal number of servers.
- If switch 2 is the master and it has one active server fewer than switch 1, then switch 1 becomes the master.

The user can implement this behavior by configuring the switch for tracking as follows:

1. Set the priority for switch 1 to 101.
2. Leave the priority for switch 2 at the default value of 100.
3. On both switches, enable tracking based on ports (**ports**), interfaces (**ifs**), or virtual routers (**vr**). You can choose any combination of tracking parameters, based on your network configuration.

   Note: There is no shortcut to setting tracking parameters. The goals must first be set and the outcomes of various configurations and scenarios analyzed to find settings that meet the goals.

High Availability Configurations

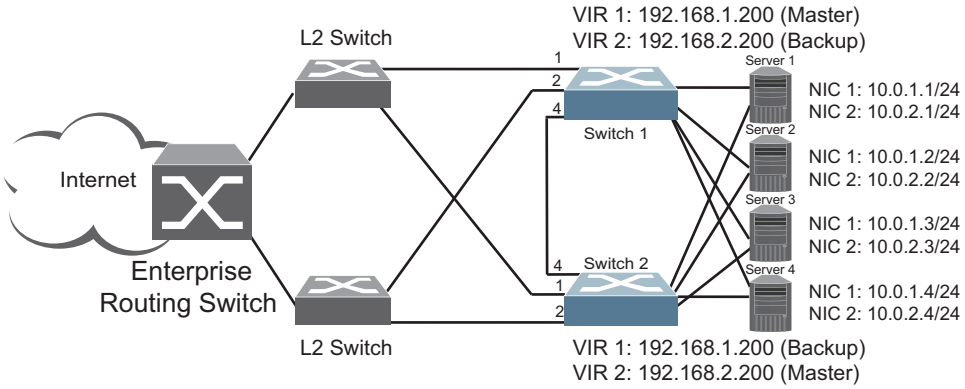
CN4093s offer flexibility in implementing redundant configurations. This section discusses the more useful and easily deployed configurations:

- "Active-Active Configuration" on page 535
- "Hot-Standby Configuration" on page 539

Active-Active Configuration

Figure 63 shows an example configuration where two CN4093s are used as VRRP routers in an active-active configuration. In this configuration, both switches respond to packets.

Figure 63. Active-Active High-Availability Configuration



Although this example shows only two switches, there is no limit on the number of switches used in a redundant configuration. It is possible to implement an active-active configuration across all the VRRP-capable switches in a LAN.

Each VRRP-capable switch in an active-active configuration is autonomous. Switches in a virtual router need not be identically configured.

In the scenario illustrated in Figure 63, traffic destined for IPv4 address 10.0.1.1 is forwarded through the Layer 2 switch at the top of the drawing, and ingresses CN4093 1 on port EXT1. Return traffic uses default gateway 1 (192.168.1.1).

If the link between CN4093 1 and the Layer 2 switch fails, CN4093 2 becomes the Master because it has a higher priority. Traffic is forwarded to CN4093 2, which forwards it to CN4093 1 through port EXT4. Return traffic uses default gateway 2 (192.168.2.1) and is forwarded through the Layer 2 switch at the bottom of the drawing.

To implement the active-active example, perform the following switch configuration.

Task 1: Configure CN4093 1

1. Configure client and server interfaces.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 192.168.1.100 255.255.255.0
CN 4093(config-ip-if)# vlan 10
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 192.168.2.101 255.255.255.0
CN 4093(config-ip-if)# vlan 20
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ip address 10.0.1.100 255.255.255.0
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 4
CN 4093(config-ip-if)# ip address 10.0.2.101 255.255.255.0
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
```

2. Configure the default gateways. Each default gateway points to a Layer 3 router.

```
CN 4093(config)# ip gateway 1 address 192.168.1.1
CN 4093(config)# ip gateway 1 enable
CN 4093(config)# ip gateway 2 address 192.168.2.1
CN 4093(config)# ip gateway 2 enable
```

3. Turn on VRRP and configure two Virtual Interface Routers.

```
CN 4093(config)# router vrrp
CN 4093(config-vrrp)# enable
CN 4093(config-vrrp)# virtual-router 1 virtual-router-id 1
CN 4093(config-vrrp)# virtual-router 1 interface 1
CN 4093(config-vrrp)# virtual-router 1 address 192.168.1.200
CN 4093(config-vrrp)# virtual-router 1 enable
CN 4093(config-vrrp)# virtual-router 2 virtual-router-id 2
CN 4093(config-vrrp)# virtual-router 2 interface 2
CN 4093(config-vrrp)# virtual-router 2 address 192.168.2.200
CN 4093(config-vrrp)# virtual-router 2 enable
```

4. Enable tracking on ports. Set the priority of Virtual Router 1 to 101, so that it becomes the Master.

```
CN 4093(config-vrrp)# virtual-router 1 track ports
CN 4093(config-vrrp)# virtual-router 1 priority 101
CN 4093(config-vrrp)# virtual-router 2 track ports
CN 4093(config-vrrp)# exit
```

5. Configure ports.

```
CN 4093(config)# vlan 10
CN 4093(config-vlan)# exit
CN 4093(config)# interface port EXT1
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan add 10
CN 4093(config-if)# exit

CN 4093(config)# vlan 20
CN 4093(config-vlan)# exit
CN 4093(config)# interface port EXT2
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan add 20
CN 4093(config-if)# exit
```

6. Turn off Spanning Tree Protocol globally..

```
CN 4093(config)# no spanning-tree stp 1
```

Task 2: Configure CN4093 2

1. Configure client and server interfaces.

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 192.168.1.101 255.255.255.0
CN 4093(config-ip-if)# vlan 10
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 192.168.2.100 255.255.255.0
CN 4093(config-ip-if)# vlan 20
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 3
CN 4093(config-ip-if)# ip address 10.0.1.101 255.255.255.0
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 4
CN 4093(config-ip-if)# ip address 10.0.2.100 255.255.255.0
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
```

2. Configure the default gateways. Each default gateway points to a Layer 3 router.

```
CN 4093(config)# ip gateway 1 address 192.168.2.1
CN 4093(config)# ip gateway 1 enable
CN 4093(config)# ip gateway 2 address 192.168.1.1
CN 4093(config)# ip gateway 2 enable
```

3. Turn on VRRP and configure two Virtual Interface Routers.

```
CN 4093(config)# router vrrp
CN 4093(config-vrrp)# enable
CN 4093(config-vrrp)# virtual-router 1 virtual-router-id 1
CN 4093(config-vrrp)# virtual-router 1 interface 1
CN 4093(config-vrrp)# virtual-router 1 address 192.168.1.200
CN 4093(config-vrrp)# virtual-router 1 enable
CN 4093(config-vrrp)# virtual-router 2 virtual-router-id 2
CN 4093(config-vrrp)# virtual-router 2 interface 2
CN 4093(config-vrrp)# virtual-router 2 address 192.168.2.200
CN 4093(config-vrrp)# virtual-router 2 enable
```

4. Enable tracking on ports. Set the priority of Virtual Router 2 to 101, so that it becomes the Master.

```
CN 4093(config-vrrp)# virtual-router 1 track ports
CN 4093(config-vrrp)# virtual-router 2 track ports
CN 4093(config-vrrp)# virtual-router 2 priority 101
CN 4093(config-vrrp)# exit
```

5. Configure ports.

```
CN 4093(config)# vlan 10
CN 4093(config-vlan)# exit
CN 4093(config)# interface port EXT1
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan add 10
CN 4093(config-if)# exit

CN 4093(config)# vlan 20
CN 4093(config-vlan)# exit
CN 4093(config)# interface port EXT2
CN 4093(config-if)# switchport mode trunk
CN 4093(config-if)# switchport trunk allowed vlan add 20
CN 4093(config-if)# exit
```

6. Turn off Spanning Tree Protocol globally.

```
CN 4093(config)# no spanning-tree stp 1
```

Hot-Standby Configuration

The primary application for VRRP-based hot-standby is to support Network Adapter Teaming on your server blades. With Network Adapter Teaming, the NICs on each server share the same IPv4 address, and are configured into a team. One NIC is the primary link, and the others are backup links. For more details, refer to the NetXen 10 Gb Ethernet Adapter documentation.

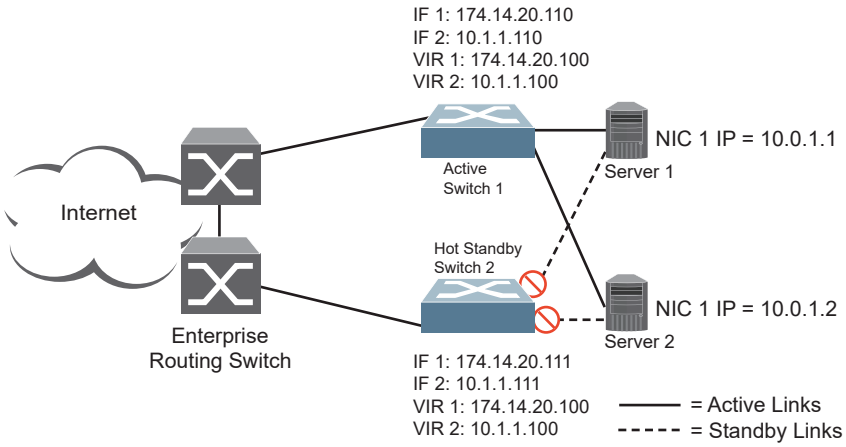
A hot-standby configuration allows all processes to failover to a standby switch if any type of failure should occur. All Virtual Interface Routers (VIRs) are bundled into one Virtual Router group, and then they failover together. When there is a failure that causes the VRRP Master to failover to the Standby, then the original primary switch temporarily disables the internal server links, which, in turn, causes the NIC teams to failover as well.

Note: When using hot-standby redundancy, peer switches should have an equal number of connected ports.

If hot-standby is implemented in a looped environment, the hot-standby feature automatically disables the hot-standby ports on the VRRP Standby. If the Master switch should failover to the Standby switch, it would change the hot-standby ports from *disabled* to *forwarding*, without relying on Spanning Tree or manual intervention. Therefore, Spanning Tree must be disabled.

Figure 64 illustrates a common hot-standby implementation on a single blade server. Notice that the blade server NICs are configured into a team that shares the same IPv4 address across both NICs. Because only one link can be active at a time, the hot-standby feature controls the NIC failover by having the Standby switch disable its internal ports (holding down the server links).

Figure 64. Hot-Standby Configuration



Task 1: Configure CN4093 1

1. On CN4093 1, configure the interfaces for clients (174.14.20.110) and servers (10.1.1.110).

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 174.14.20.110 (Define IPv4 address for interface 1)
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 10.1.1.110     (Define IPv4 address for interface 2)
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
```

2. Configure Virtual Interface Routers.

```
CN 4093(config)# router vrrp
CN 4093(config-vrrp)# enable
CN 4093(config-vrrp)# virtual-router 1 virtual-router-id 1
CN 4093(config-vrrp)# virtual-router 1 interface 1
CN 4093(config-vrrp)# virtual-router 1 address 174.14.20.100
CN 4093(config-vrrp)# virtual-router 1 enable
CN 4093(config-vrrp)# virtual-router 2 virtual-router-id 2
CN 4093(config-vrrp)# virtual-router 2 interface 2
CN 4093(config-vrrp)# virtual-router 2 address 10.1.1.100
CN 4093(config-vrrp)# virtual-router 2 enable
```

3. Enable VRRP Hot Standby.

```
CN 4093(config-vrrp)# hot-standby      (Enable Hot Standby)
```

4. Configure VRRP Group parameters. Set the VRRP priority to 101, so that this switch is the Master.

```
CN 4093(config-vrrp)# group enable     (Enable Virtual Router Group)
CN 4093(config-vrrp)# group virtual-router-id 1  (Set Virtual Router ID for Group)
CN 4093(config-vrrp)# group interface 1    (Set interface for Group)
CN 4093(config-vrrp)# group priority 101   (Set VRRP priority to 101)
CN 4093(config-vrrp)# group track ports    (Enable tracking on ports)
```

5. Turn off Spanning Tree Protocol globally..

```
CN 4093(config)# no spanning-tree stp 1
```

Task 2: Configure CN4093 2

1. On CN4093 2, configure the interfaces for clients (174.14.20.111) and servers (10.1.1.111).

```
CN 4093(config)# interface ip 1
CN 4093(config-ip-if)# ip address 174.14.20.111 (Define IPv4 address for interface 1)
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit

CN 4093(config)# interface ip 2
CN 4093(config-ip-if)# ip address 10.1.1.111     (Define IPv4 address for interface 2)
CN 4093(config-ip-if)# enable
CN 4093(config-ip-if)# exit
```

2. Configure Virtual Interface Routers.

```
CN 4093(config)# router vrrp
CN 4093(config-vrrp)# enable
CN 4093(config-vrrp)# virtual-router 1 virtual-router-id 1
CN 4093(config-vrrp)# virtual-router 1 interface 1
CN 4093(config-vrrp)# virtual-router 1 address 174.14.20.100
CN 4093(config-vrrp)# virtual-router 1 enable
CN 4093(config-vrrp)# virtual-router 2 virtual-router-id 2
CN 4093(config-vrrp)# virtual-router 2 interface 2
CN 4093(config-vrrp)# virtual-router 2 address 10.1.1.100
CN 4093(config-vrrp)# virtual-router 2 enable
```

3. Enable VRRP Hot Standby.

```
CN 4093(config-vrrp)# hot-standby
```

4. Configure VRRP Group parameters. Use the default VRRP priority of 100, so that this switch is the Standby.

```
CN 4093(config-vrrp)# group enable      (Enable Virtual Router Group)
CN 4093(config-vrrp)# group virtual-router-id 1  (Set Virtual Router ID for Group)
CN 4093(config-vrrp)# group interface 1    (Set interface for Group)
CN 4093(config-vrrp)# group track ports    (Enable tracking on ports)
```

5. Turn off Spanning Tree Protocol globally. .

```
CN 4093(config)# spanning-tree mode disable
```

Part 7: Network Management

Chapter 35. Link Layer Discovery Protocol

The Enterprise NOS software support Link Layer Discovery Protocol (LLDP). This chapter discusses the use and configuration of LLDP on the switch:

LLDP Overview

Link Layer Discovery Protocol (LLDP) is an IEEE 802.1AB-2005 standard for discovering and managing network devices. LLDP uses Layer 2 (the data link layer), and allows network management applications to extend their awareness of the network by discovering devices that are direct neighbors of already known devices.

With LLDP, the CN4093 can advertise the presence of its ports, their major capabilities, and their current status to other LLDP stations in the same LAN. LLDP transmissions occur on ports at regular intervals or whenever there is a relevant change to their status. The switch can also receive LLDP information advertised from adjacent LLDP-capable network devices.

In addition to discovery of network resources, and notification of network changes, LLDP can help administrators quickly recognize a variety of common network configuration problems, such as unintended VLAN exclusions or mis-matched port aggregation membership.

The LLDP transmit function and receive function can be independently configured on a per-port basis. The administrator can allow any given port to transmit only, receive only, or both transmit and receive LLDP information.

The LLDP information to be distributed by the CN4093 ports, and that which has been collected from other LLDP stations, is stored in the switch's Management Information Base (MIB). Network Management Systems (NMS) can use Simple Network Management Protocol (SNMP) to access this MIB information. LLDP-related MIB information is read-only.

Changes, either to the local switch LLDP information or to the remotely received LLDP information, are flagged within the MIB for convenient tracking by SNMP-based management systems.

For LLDP to provide expected benefits, all network devices that support LLDP should be consistent in their LLDP configuration.

LLDP - Stacking Mode

In stacking mode, LLDP can be configured only on the ports that are not used to create the stack. The LLDP configuration menus on the stacking ports are disabled.

When configuring LLDP on a port, use the correct port syntax. See example of port syntax on page 245.

Enabling or Disabling LLDP

Global LLDP Setting

By default, LLDP is enabled on the CN4093. To turn LLDP off or on, use the following command:

```
CN 4093(config)# [no] lldp enable       (Turn LLDP on or off globally)
```

Transmit and Receive Control

The CN4093 can also be configured to transmit or receive LLDP information on a port-by-port basis. By default, when LLDP is globally enabled on the switch, CN4093 ports transmit and receive LLDP information (see the `tx_rx` option below). To change the LLDP transmit and receive state, the following commands are available:

```
CN 4093(config)# interface port <x>      (Select a switch port)
CN 4093(config-if)# lldp admin-status tx_rx  (Transmit and receive LLDP)
CN 4093(config-if)# lldp admin-status tx_only  (Only transmit LLDP)
CN 4093(config-if)# lldp admin-status rx_only  (Only receive LLDP)
CN 4093(config-if)# no lldp admin-status      (Do not participate in LLDP)
CN 4093(config-if)# exit                 (Exit port mode)
```

To view the LLDP transmit and receive status, use the following commands:

```
CN 4093(config)# show lldp port          (status of all ports)
CN 4093(config)# show interface port <n> lldp  (status of selected port)
```

LLDP Transmit Features

Numerous LLDP transmit options are available, including scheduled and minimum transmit interval, expiration on remote systems, SNMP trap notification, and the types of information permitted to be shared.

Note: In stacking mode, only the stack Master transmits LLDP information for all the ports in a stack. The stack MAC address is used as the source address in the LLDP packets.

Scheduled Interval

The CN4093 can be configured to transmit LLDP information to neighboring devices once each 5 to 32768 seconds. The scheduled interval is global; the same interval value applies to all LLDP transmit-enabled ports. However, to help balance LLDP transmissions and keep them from being sent simultaneously on all ports, each port maintains its own interval clock, based on its own initialization or reset time. This allows switch-wide LLDP transmissions to be spread out over time, though individual ports comply with the configured interval.

The global transmit interval can be configured using the following command:

```
CN 4093(config)# lldp refresh-interval <interval>
```

where *interval* is the number of seconds between LLDP transmissions. The range is 5 to 32768. The default is 30 seconds.

Minimum Interval

In addition to sending LLDP information at scheduled intervals, LLDP information is also sent when the CN4093 detects relevant changes to its configuration or status (such as when ports are enabled or disabled). To prevent the CN4093 from sending multiple LLDP packets in rapid succession when port status is in flux, a transmit delay timer can be configured.

The transmit delay timer represents the minimum time permitted between successive LLDP transmissions on a port. Any interval-driven or change-driven updates will be consolidated until the configured transmit delay expires.

The minimum transmit interval can be configured using the following command:

```
CN 4093(config)# lldp transmission-delay <interval>
```

where *interval* is the minimum number of seconds permitted between successive LLDP transmissions on any port. The range is 1 to one-quarter of the scheduled transmit interval (**`lldp refresh-interval`** *<value>*), up to 8192. The default is 2 seconds.

Time-to-Live for Transmitted Information

The transmitted LLDP information is held by remote systems for a limited time. A time-to-live parameter allows the switch to determine how long the transmitted data should be held before it expires. The hold time is configured as a multiple of the configured transmission interval.

```
CN 4093(config)# lldp holdtime-multiplier <multiplier>
```

where *multiplier* is a value between 2 and 10. The default value is 4, meaning that remote systems will hold the port's LLDP information for 4 **x** the 30-second `msgtxint` value, or 120 seconds, before removing it from their MIB.

Trap Notifications

If SNMP is enabled on the CN4093 (see "Using Simple Network Management Protocol" on page 39), each port can be configured to send SNMP trap notifications whenever LLDP transmissions are sent. By default, trap notification is disabled for each port. The trap notification state can be changed using the following commands:

```
CN 4093(config)# interface port <x>
CN 4093(config-if)# [no] lldp trap-notification
CN 4093(config-if)# exit
```

In addition to sending LLDP information at scheduled intervals, LLDP information is also sent when the CN4093 detects relevant changes to its configuration or status (such as when ports are enabled or disabled). To prevent the CN4093 from sending multiple trap notifications in rapid succession when port status is in flux, a global trap delay timer can be configured.

The trap delay timer represents the minimum time permitted between successive trap notifications on any port. Any interval-driven or change-driven trap notices from the port will be consolidated until the configured trap delay expires.

The minimum trap notification interval can be configured using the following command:

```
CN 4093(config)# lldp trap-notification-interval <interval>
```

where *interval* is the minimum number of seconds permitted between successive LLDP transmissions on any port. The range is 1 to 3600. The default is 5 seconds.

If SNMP trap notification is enabled, the notification messages can also appear in the system log. This is enabled by default. To change whether the SNMP trap notifications for LLDP events appear in the system log, use the following commands:

```
CN 4093(config)# [no] logging log lldp
```

Changing the LLDP Transmit State

When the port is disabled, or when LLDP transmit is turned off for the port using the `admstat` command's `rx_only` or `disabled` options (see "Transmit and Receive Control" on page 547), a final LLDP packet is transmitted with a time-to-live value of 0. Neighbors that receive this packet will remove the LLDP information associated with the CN4093 port from their MIB.

In addition, if LLDP is fully disabled on a port (using `admstat disabled`) and later re-enabled, the CN4093 will temporarily delay resuming LLDP transmissions on the port in order to allow the port LLDP information to stabilize. The reinitialization delay interval can be globally configured for all ports using the following command:

```
CN 4093(config)# lldp reinit-delay <interval>
```

where *interval* is the number of seconds to wait before resuming LLDP transmissions. The range is between 1 and 10. The default is 2 seconds.

Types of Information Transmitted

When LLDP transmission is permitted on the port (see "Enabling or Disabling LLDP" on page 547), the port advertises the following required information in type/length/value (TLV) format:

- Chassis ID
- Port ID
- LLDP Time-to-Live

LLDP transmissions can also be configured to enable or disable inclusion of optional information, using the following command:

```
CN 4093(config)# interface port <x>
CN 4093(config-if)# [no] lldp tlv <type>
CN 4093(config-if)# exit
```

where *type* is an LLDP information option from Table 39:

Table 39. *LLDP Optional Information Types*

Type	Description	Default
`portdesc`	Port Description	Enabled
`sysname`	System Name	Enabled
`sysdescr`	System Description	Enabled
`syscap`	System Capabilities	Enabled
`mgmtaddr`	Management Address	Enabled
`portvid`	IEEE 802.1 Port VLAN ID	Disabled
`portprot`	IEEE 802.1 Port and Protocol VLAN ID	Disabled
`vlanname`	IEEE 802.1 VLAN Name	Disabled
`protid`	IEEE 802.1 Protocol Identity	Disabled
`macphy`	IEEE 802.3 MAC/PHY Configuration/Status, including the auto-negotiation, duplex, and speed status of the port.	Disabled
`powermdi`	IEEE 802.3 Power via MDI, indicating the capabilities and status of devices that require or provide power over twisted-pair copper links.	Disabled
`linkaggr`	IEEE 802.3 Link Aggregation status for the port.	Disabled
`framesz`	IEEE 802.3 Maximum Frame Size for the port.	Disabled

Table 39. *LLDP Optional Information Types (continued)*

Type	Description	Default
`dcbx`	Data Center Bridging Capability Exchange Protocol (DCBX) for the port.	Enabled
`all`	Select all optional LLDP information for inclusion or exclusion.	Disabled

LLDP Receive Features

Types of Information Received

When the LLDP receive option is enabled on a port (see "Enabling or Disabling LLDP" on page 547), the port may receive the following information from LLDP-capable remote systems:

- Chassis Information
- Port Information
- LLDP Time-to-Live
- Port Description
- System Name
- System Description
- System Capabilities Supported/Enabled
- Remote Management Address

The CN4093 stores the collected LLDP information in the MIB. Each remote LLDP-capable device is responsible for transmitting regular LLDP updates. If the received updates contain LLDP information changes (to port state, configuration, LLDP MIB structures, deletion), the switch will set a change flag within the MIB for convenient notification to SNMP-based management systems.

Note: In stacking mode, both the Master and the Backup receive LLDP information for all the ports in a stack and update the LLDP table. The Master and Backup switches synchronize the LLDP tables.

Viewing Remote Device Information

LLDP information collected from neighboring systems can be viewed in numerous ways:

- Using a centrally-connected LLDP analysis server
- Using an SNMP agent to examine the CN4093 MIB
- Using the CN4093 Browser-Based Interface (BBI)
- Using CLI or isCLI commands on the CN4093

Using the isCLI the following command displays remote LLDP information:

```
CN 4093(config)# show lldp remote-device [<index number>]
```

To view a summary of remote information, omit the *Index number* parameter. For example:

```
CN 4093(config)# show lldp remote-device
LLDP Remote Devices Information
Legend(possible values in DMAC column) :
NB - Nearest Bridge - 01-80-C2-00-00-0E
NnTB - Nearest non-TPMR Bridge - 01-80-C2-00-00-03
NCB - Nearest Customer Bridge - 01-80-C2-00-00-00
Total number of current entries: 1
LocalPort |Index |Remote Chassis ID |Remote Port |Remote System Name|DMAC
----------|------|-----------------|------------|------------------|----
EXT3      | 1    |00 18 b1 33 1d 00 |    23      |    C12           | NB
```

To view detailed information for a remote device, specify the *Index number* as found in the summary. For example, in keeping with the sample summary, to list details for the first remote device (with an `Index` value of 1), use the following command:

```
CN 4093(config)# show lldp remote-device 1
Local Port Alias: EXT3
        Remote Device Index     : 1
        Remote Device TTL       : 99
        Remote Device RxChanges : false
        Chassis Type            : Mac Address
        Chassis Id              : 00-18-b1-33-1d-00
        Port Type               : Locally Assigned
        Port Id                 : 23
        Port Description        : EXT7

        System Name        :
        System Description : Lenovo Flex System Fabric CN4093 10 Gb
Converged Scalable Switch, Enterprise NOS: version 8.4, boot image:
version 6.9.1.14

        System Capabilities Supported : bridge, router
        System Capabilities Enabled   : bridge, router

        Remote Management Address:
                Subtype             : IPv4
                Address             : 10.100.120.181
                Interface Subtype   : ifIndex
                Interface Number    : 128
                Object Identifier   :
```

Note: Received LLDP information can change very quickly. When using show commands, it is possible that flags for some expected events may be too short-lived to be observed in the output.

To view detailed information of all remote devices, use the following command:

```
CN 4093(config)# show lldp remote-device detail
Local Port Alias: EXT22
        Remote Device Index     : 1
        Remote Device TTL       : 94
        Remote Device RxChanges : false
        Chassis Type            : Mac Address
        Chassis Id              : 74-99-75-74-c5-00
        Port Type               : Locally Assigned
        Port Id                 : 42
        Port Description        : 42

        System Name        : GFC
        System Description : Lenovo RackSwitch G8264CS, Lenovo
Networking OS: version 7.8.0.43, Boot image: version 7.8.0.43
        System Capabilities Supported : bridge, router
        System Capabilities Enabled   : bridge, router

        Remote Management Address:
                Subtype             : IPv4
                Address             : 11.1.58.5
                Interface Subtype   : ifIndex
                Interface Number    : 58
                Object Identifier   :

Local Port Alias: EXT24
        Remote Device Index     : 2
        Remote Device TTL       : 108
        Remote Device RxChanges : false
        Chassis Type            : Mac Address
        Chassis Id              : 74-99-75-1c-71-00
        Port Type               : Locally Assigned
        Port Id                 : 56
        Port Description        : EXT14

        System Name        : CFC
        System Description : Lenovo Flex System CN4093 10Gb Converged
Scalable Switch, Lenovo Networking OS: version 7.8.0.48, Boot image:
version 7.8.0.48
        System Capabilities Supported : bridge, router
        System Capabilities Enabled   : bridge, router

        Remote Management Address:
                Subtype             : IPv4
                Address             : 11.1.78.7
                Interface Subtype   : ifIndex
                Interface Number    : 78
                Object Identifier   :
```

Time-to-Live for Received Information

Each remote device LLDP packet includes an expiration time. If the switch port does not receive an LLDP update from the remote device before the time-to-live clock expires, the switch will consider the remote information to be invalid, and will remove all associated information from the MIB.

Remote devices can also intentionally set their LLDP time-to-live to 0, indicating to the switch that the LLDP information is invalid and should be immediately removed.

LLDP Example Configuration

1. Turn LLDP on globally.

```
CN 4093(config)# lldp enable
```

2. Set the global LLDP timer features.

```
CN 4093(config)# lldp refresh-interval 30   (Transmit each 30 seconds)
CN 4093(config)# lldp transmission-delay 2  (No more often than 2 sec.)
CN 4093(config)# lldp holdtime-multiplier 4  (Remote hold 4 intervals)
CN 4093(config)# lldp reinit-delay 2  (Wait 2 sec. after reinit.)
CN 4093(config)# lldp trap-notification-interval 5  (Minimum 5 sec. between)
```

3. Set LLDP options for each port.

```
CN 4093(config)# interface port <n>      (Select a switch port)
CN 4093(config-if)# lldp admin-status tx_rx  (Transmit and receive LLDP)
CN 4093(config-if)# lldp trap-notification   (Enable SNMP trap notifications)
CN 4093(config-if)# lldp tlv all       (Transmit all optional information)
CN 4093(config-if)# exit
```

4. Enable syslog reporting.

```
CN 4093(config)# logging log lldp
```

5. Verify the configuration settings:

```
CN 4093(config)# show lldp
```

6. View remote device information as needed.

```
CN 4093(config)# show lldp remote-device
   or
CN 4093(config)# show lldp remote-device <index number>
   or
CN 4093(config)# show lldp remote-devices detail
```

Chapter 36. Simple Network Management Protocol

Enterprise NOS provides Simple Network Management Protocol (SNMP) version 1, version 2, and version 3 support for access through any network management software, such as Lenovo Director.

SNMP Version 1

To access the SNMP agent on the CN4093, the read and write community strings on the SNMP manager should be configured to match those on the switch.

The read and write community strings on the switch can be changed using the following commands:

```
CN 4093(config)# snmp-server read-community <1-32 characters>
   -and-
CN 4093(config)# snmp-server write-community <1-32 characters>
```

The SNMP manager should be able to reach the management interface or any one of the IP interfaces on the switch.

For the SNMP manager to receive the SNMPv1 traps sent out by the SNMP agent on the switch, configure the trap host on the switch with the following command:

```
CN 4093(config)# snmp-server trap-source <trap source IP interface>
CN 4093(config)# snmp-server host <IPv4 address> <trap host community string>
```

Note: You can use a loopback interface to set the source IP address for SNMP traps. Use the following command to apply a configured loopback interface:
`CN 4093(config)# snmp-server trap-source loopback <1-5>`

SNMP Version 3

SNMP version 3 (SNMPv3) is an enhanced version of the Simple Network Management Protocol, approved by the Internet Engineering Steering Group in March, 2002. SNMPv3 contains additional security and authentication features that provide data origin authentication, data integrity checks, timeliness indicators and encryption to protect against threats such as masquerade, modification of information, message stream modification and disclosure.

SNMPv3 allows clients to query the MIBs securely.

SNMPv3 configuration is managed using the following command path:

```
CN 4093(config)# snmp-server ?
```

For more information on SNMP MIBs and the commands used to configure SNMP on the switch, see the *Enterprise NOS 8.4 Command Reference*.

Default Configuration

Enterprise NOS has SNMPv3 disabled by default. If a user-created SNMPv3 user is found on the system, SNMPv3 is enabled for backwards compatibility.

Up to 17 SNMP users can be configured on the switch. To modify an SNMP user, enter the following commands:

```
CN 4093(config)# snmp-server user <1-17> name <1-32 characters>
```

Users can be configured to use the authentication/privacy options. The CN4093 support two authentication algorithms: MD5 and SHA, as specified in the following command:

```
CN 4093(config)# snmp-server user <1-17> authentication-protocol {md5|sha}
authentication-password
   -or-
CN 4093(config)# snmp-server user <1-17> authentication-protocol none
```

User Configuration Example

1. To configure a user with name "admin," authentication type MD5, and authentication password of "admin," privacy option DES with privacy password of "admin," use the following ISCLI commands.

```
CN 4093(config)# snmp-server user 5 name admin
CN 4093(config)# snmp-server user 5 authentication-protocol md5
authentication-password
Changing authentication password; validation required:
Enter current admin password:          <admin. password>
Enter new authentication password:     <auth. password>
Re-enter new authentication password: <auth. password>
New authentication password accepted.

CN 4093(config)# snmp-server user 5 privacy-protocol des privacy-password
Changing privacy password; validation required:
Enter current admin password:          <admin. password>
Enter new privacy password:            <privacy password>
Re-enter new privacy password:         <privacy password>
New privacy password accepted.
```

2. Configure a user access group, along with the views the group may access. Use the access table to configure the group's access level.

```
CN 4093(config)# snmp-server access 5 name admingrp
CN 4093(config)# snmp-server access 5 level authpriv
CN 4093(config)# snmp-server access 5 read-view iso
CN 4093(config)# snmp-server access 5 write-view iso
CN 4093(config)# snmp-server access 5 notify-view iso
```

   Because the read view, write view, and notify view are all set to "iso," the user type has access to all private and public MIBs.

3. Assign the user to the user group. Use the group table to link the user to a particular access group.

```
CN 4093(config)# snmp-server group 5 user-name admin
CN 4093(config)# snmp-server group 5 group-name admingrp
```

   If you want to allow user access only to certain MIBs, see "View-Based Configuration," next.

View-Based Configurations

- Switch `User` equivalent

  To configure an SNMP user equivalent to the switch "user" login, use the following configuration:

```
(Configure the user)
CN 4093(config)# snmp-server user 4 name usr
(Configure access group 3)
CN 4093(config)# snmp-server access 3 name usrgrp
CN 4093(config)# snmp-server access 3 read-view usr
CN 4093(config)# snmp-server access 3 write-view usr
CN 4093(config)# snmp-server access 3 notify-view usr
(Assign user to access group 3)
CN 4093(config)# snmp-server group 3 user-name usr
CN 4093(config)# snmp-server group 3 group-name usrgrp
(Create views for user)
CN 4093(config)# snmp-server view 6 name usr
CN 4093(config)# snmp-server view 6 tree 1.3.6.1.4.1.1872.2.5.1.2
(Agent information)
CN 4093(config)# snmp-server view 7 name usr
CN 4093(config)# snmp-server view 7 tree 1.3.6.1.4.1.1872.2.5.1.3
(L2 statistics)
CN 4093(config)# snmp-server view 8 name usr
CN 4093(config)# snmp-server view 8 tree 1.3.6.1.4.1.1872.2.5.2.2
(L2 information)
CN 4093(config)# snmp-server view 9 name usr
CN 4093(config)# snmp-server view 9 tree 1.3.6.1.4.1.1872.2.5.2.3
(L3 statistics)
CN 4093(config)# snmp-server view 10 name usr
CN 4093(config)# snmp-server view 10 tree 1.3.6.1.4.1.1872.2.5.2.3
(L3 information)
CN 4093(config)# snmp-server view 11 name usr
CN 4093(config)# snmp-server view 11 tree 1.3.6.1.4.1.1872.2.5.3.3
```

- Switch `Oper` equivalent

```
(Configure the user)
CN 4093(config)# snmp-server user 5 name usr
(Configure access group 3)
CN 4093(config)# snmp-server access 4 name opergrp
CN 4093(config)# snmp-server access 4 read-view oper
CN 4093(config)# snmp-server access 4 write-view oper
CN 4093(config)# snmp-server access 4 notify-view oper
(Assign oper to access group 4)
CN 4093(config)# snmp-server group 4 user-name oper
CN 4093(config)# snmp-server group 4 group-name opergrp
(Create views for oper)
CN 4093(config)# snmp-server view 20 name oper
CN 4093(config)# snmp-server view 20 tree 1.3.6.1.4.1.1872.2.5.1.2
(Agent information)
CN 4093(config)# snmp-server view 21 name oper
CN 4093(config)# snmp-server view 21 tree 1.3.6.1.4.1.1872.2.5.1.3
(L2 statistics)
CN 4093(config)# snmp-server view 22 name oper
CN 4093(config)# snmp-server view 22 tree 1.3.6.1.4.1.1872.2.5.2.2
(L2 information)
CN 4093(config)# snmp-server view 23 name oper
CN 4093(config)# snmp-server view 23 tree 1.3.6.1.4.1.1872.2.5.2.3
(L3 statistics)
CN 4093(config)# snmp-server view 24 name oper
CN 4093(config)# snmp-server view 24 tree 1.3.6.1.4.1.1872.2.5.2.3
(L3 information)
CN 4093(config)# snmp-server view 25 name oper
CN 4093(config)# snmp-server view 25 tree 1.3.6.1.4.1.1872.2.5.3.3
```

Secure Audit Logging

Flex System managers may use the authentication and encryption protocols of SNMPv3 to securely audit the switch. The audit logs record activity and severity for the overall system, user, and application processes. These logs can be used to trace a user's actions, monitor switch alerts, and confirm intrusion detection.

Networking OS uses SNMPv3 authorization to forward the logs securely to the management tool via the chassis management module (CMM). The switch supports both retrieving the logs via SNMP 'Get' requests and the forwarding of event logs via SNMP traps. Supported management tools are xHMC and other (security and information event management) SIEM tools like Qradar.

Security audit logging refers to the following event types:

- NTP Server/DHCP server configuration changes
- Switch management IP address changes
- OSPF/BGP/RIP authentication changes
- Software Resource alert :ARP Table/IP table/Route table/OSPF table full
- L3 Link down/up

Note: Audit logging is enabled by default and cannot be disabled. The audit logs are accessed remotely via SNMPv3 hosts.

Use the following commands to locally manage the logs:

```
CN 4093(config)# show sal reverse          (Display most recent logs first)
CN 4093(config)# clear sal                 (Clear audit logs)
```

Configuring SNMP Trap Hosts

Follow these instructions to configure SNMP trap hosts.

SNMPv1 Trap Host Configuration

1. Configure a user with no authentication and password.

```
CN 4093(config)# snmp-server user 10 name v1trap
```

2. Configure an access group and group table entries for the user. Use the following menu to specify which traps can be received by the user:

```
CN 4093(config)# snmp-server access <user number>
```

In the following example the user will receive the traps sent by the switch.

```
CN 4093(config)# snmp-server access 10   (Access group to view SNMPv1 traps)
     name v1trap
     security snmpv1
     notify-view iso
CN 4093(config)# snmp-server group 10     (Assign user to the access group)
     security snmpv1
     user-name v1trap
     group-name v1trap
```

3. Configure an entry in the notify table.

```
CN 4093(config)# snmp-server notify 10 name v1trap
CN 4093(config)# snmp-server notify 10 tag v1trap
```

4. Specify the IPv4 address and other trap parameters in the `targetAddr` and `targetParam` tables. Use the following commands to specify the user name associated with the targetParam table:

```
CN 4093(config)# snmp-server target-address 10 name v1trap address
10.70.70.190
CN 4093(config)# snmp-server target-address 10 parameters-name v1param
CN 4093(config)# snmp-server target-address 10 taglist v1param
CN 4093(config)# snmp-server target-parameters 10 name v1param
CN 4093(config)# snmp-server target-parameters 10 user-name v1only
CN 4093(config)# snmp-server target-parameters 10 message snmpv1
```

Note: Enterprise NOS 8.4 supports only IPv4 addresses for SNMP trap hosts.

5. Use the community table to specify which community string is used in the trap.

```
CN 4093(config)# snmp-server community 10(Define the community string)
     index v1trap
     name public
     user-name v1trap
```

SNMPv2 Trap Host Configuration

The SNMPv2 trap host configuration is similar to the SNMPv1 trap host configuration. Wherever you specify the model, use `snmpv2` instead of `snmpv1`.

```
CN 4093(config)# snmp-server user 10 name v2trap

CN 4093(config)# snmp-server group 10 security snmpv2
CN 4093(config)# snmp-server group 10 user-name v2trap
CN 4093(config)# snmp-server group 10 group-name v2trap
CN 4093(config)# snmp-server access 10 name v2trap
CN 4093(config)# snmp-server access 10 security snmpv2
CN 4093(config)# snmp-server access 10 notify-view iso

CN 4093(config)# snmp-server notify 10 name v2trap
CN 4093(config)# snmp-server notify 10 tag v2trap

CN 4093(config)# snmp-server target-address 10 name v2trap
address 100.10.2.1
CN 4093(config)# snmp-server target-address 10 taglist v2trap
CN 4093(config)# snmp-server target-address 10 parameters-name
v2param
CN 4093(config)# snmp-server target-parameters 10 name v2param
CN 4093(config)# snmp-server target-parameters 10 message snmpv2c
CN 4093(config)# snmp-server target-parameters 10 user-name v2trap
CN 4093(config)# snmp-server target-parameters 10 security snmpv2

CN 4093(config)# snmp-server community 10 index v2trap
CN 4093(config)# snmp-server community 10 user-name v2trap
```

Note: Enterprise NOS 8.4 supports only IPv4 addresses for SNMPv1 and SNMP v2 trap hosts.

SNMPv3 Trap Host Configuration

To configure a user for SNMPv3 traps, you can choose to send the traps with both privacy and authentication, with authentication only, or without privacy or authentication.

This is configured in the access table using the following commands:

```
CN 4093(config)# snmp-server access <1-32> level
CN 4093(config)# snmp-server target-parameters <1-16>
```

Configure the user in the user table accordingly.

It is not necessary to configure the community table for SNMPv3 traps because the community string is not used by SNMPv3.

The following example shows how to configure a SNMPv3 user `v3trap` with authentication only:

```
CN 4093(config)# snmp-server user 11 name v3trap
CN 4093(config)# snmp-server user 11 authentication-protocol md5
authentication-password
Changing authentication password; validation required:
Enter current admin password:         <admin. password>
Enter new authentication password:    <auth. password>
Re-enter new authentication password: <auth. password>
New authentication password accepted.
CN 4093(config)# snmp-server access 11 notify-view iso
CN 4093(config)# snmp-server access 11 level authnopriv
CN 4093(config)# snmp-server group 11 user-name v3trap
CN 4093(config)# snmp-server group 11 tag v3trap
CN 4093(config)# snmp-server notify 11 name v3trap
CN 4093(config)# snmp-server notify 11 tag v3trap
CN 4093(config)# snmp-server target-address 11 name v3trap address
47.81.25.66
CN 4093(config)# snmp-server target-address 11 taglist v3trap
CN 4093(config)# snmp-server target-address 11 parameters-name v3param
CN 4093(config)# snmp-server target-parameters 11 name v3param
CN 4093(config)# snmp-server target-parameters 11 user-name v3trap
CN 4093(config)# snmp-server target-parameters 11 level authNoPriv
```

SNMP MIBs

The Enterprise NOS SNMP agent supports SNMP version 3. Security is provided through SNMP community strings. The default community strings are "`public`" for `SNMP GET` operation and "`private`" for `SNMP SET` operation. The community string can be modified only through the Command Line Interface (CLI). Detailed SNMP MIBs and trap definitions of the Enterprise NOS SNMP agent are contained in the following Enterprise NOS enterprise MIB document:

```
GbScSE-10G-L2L3.mib
```

The Enterprise NOS SNMP agent supports the following standard MIBs:

- `dot1x.mib`
- `ieee8021ab.mib`
- `ieee8023ad.mib`
- `lldpxdcbx.mib`
- `rfc1213.mib`
- `rfc1215.mib`
- `rfc1493.mib`
- `rfc1573.mib`
- `rfc1643.mib`
- `rfc1657.mib`
- `rfc1757.mib`
- `rfc1850.mib`
- `rfc1907.mib`
- `rfc2037.mib`
- `rfc2233.mib`
- `rfc2465.mib`
- `rfc2571.mib`
- `rfc2572.mib`
- `rfc2573.mib`
- `rfc2574.mib`
- `rfc2575.mib`
- `rfc2576.mib`
- `rfc3176.mib`

The following Fibre Channel/FCoE MIBs are supported:

- `ibm-nos-fc-fcoe.mib`
- `fc_fe_exp.mib`
- `fcmgmt.mib`
- `stack_rfc2837.mib`

The Enterprise NOS SNMP agent supports the following generic traps as defined in RFC 1215:

- ColdStart
- WarmStart
- LinkDown
- LinkUp
- AuthenticationFailure

The SNMP agent also supports two Spanning Tree traps as defined in RFC 1493:

- NewRoot
- TopologyChange

The following are the enterprise SNMP traps supported in Enterprise NOS:

Table 40. *Enterprise NOS-Supported Enterprise SNMP Traps*

Trap Name	Description
`altSwLoginFailure`	Signifies that someone failed to enter a valid username/password combination. `altSwTrapDisplayString` specifies whether the login attempt was from CONSOLE or TELNET. In case of TELNET login it also specifies the IP address of the host from which the attempt was made.
`altSwValidLogin`	Signifies that a user login has occurred.
`altSwApplyComplete`	Signifies that new configuration has been applied.
`altSwSaveComplete`	Signifies that new configuration has been saved.
`altSwFwDownloadSucess`	Signifies that firmware has been downloaded to [image1|image2|boot image].
`altSwFwDownloadFailure`	Signifies that firmware downloaded failed to [image1|image2|boot image].
`altSwValidLogout`	Signifies that a user logout has occurred.
`altSwDefAdminDisable`	Signifies that the default admin account has been disabled.
`altSwAcntStrngPswdNotMet`	Signifies that the configured password does not match strong password complexity.

Table 40. *Enterprise NOS-Supported Enterprise SNMP Traps (continued)*

Trap Name	Description
altSwAcntLocked	Signifies that account has been locked.
altSwAcntUnlocked	Signifies that account has been unlocked.
altSwSNMPBlockIPTrap	Signifies that SNMP requests are blocked; trap includes the blocked IP address.
altSwStgNewRoot	Signifies that the bridge has become the new root of the STG.
altSwCistNewRoot	Signifies that the bridge has become the new root of the CIST.
altSwStgTopologyChanged	Signifies that there was a STG topology change.
altSwCistTopologyChanged	Signifies that there was a CIST topology change.
altSwHotlinksMasterUp	Signifies that the Master interface is active.
altSwHotlinksMasterDn	Signifies that the Master interface is not active.
altSwHotlinksBackupUp	Signifies that the Backup interface is active.
altSwHotlinksBackupDn	Signifies that the Backup interface is not active.
altSwHotlinksNone	Signifies that there are no active interfaces.
altSwStgBlockingState	Signifies port state has changed to blocking state.
altSwTeamingCtrlUp	Signifies that the teaming is up.
altSwTeamingCtrlDown	Signifies that the teaming control is down.
altSwTeamingCtrlDownTearDown Blked	Signifies that the teaming control is down but teardown is blocked.
altSwTeamingCtrlError	Signifies error, action is undefined.
altSwLACPPortBlocked	Signifies that LACP is operationally down on a port, and traffic is blocked on the port.

Table 40. *Enterprise NOS-Supported Enterprise SNMP Traps (continued)*

Trap Name	Description
altSwLACPPortUnblocked	Signifies that LACP is operationally up on a port, and traffic is no longer blocked on the port.
altSwLFDPortErrdisabled	Signifies that a port is error-disabled due to excessive link flaps.
altSwVlagInstanceUp	Signifies that VLAG instance is up identified in the trap message.
altSwVlagInstanceRemoteUp	Signifies that VLAG is down but instance on the remote instance is up.
altSwVlagInstanceLocalUp	Signifies that VLAG is down but local instance is up.
altSwVlagInstanceDown	Signifies that VLAG instance is down identified in the trap message.
altSwVlagIslUp	Signifies that connection between VLAG switches is up.
altSwVlagIslDown	Signifies that connection between VLAG switches is down.
altSwDefGwUp	Signifies that the default gateway is alive. ipCurCfgGwIndex is the index of the Gateway in ipCurCfgGwTable. The range for ipCurCfgGwIndex is from 1 to ipGatewayTableMax. ipCurCfgGwAddr is the IP address of the default gateway.
altSwDefGwDown	Signifies that the default gateway is down. ipCurCfgGwIndex is the index of the Gateway in ipCurCfgGwTable. The range for ipCurCfgGwIndex is from 1 to ipGatewayTableMax. ipCurCfgGwAddr is the IP address of the default gateway.

Table 40. *Enterprise NOS-Supported Enterprise SNMP Traps (continued)*

Trap Name	Description
altSwDefGwInService	Signifies that the default gateway is up and in service. ipCurCfgGwIndex is the index of the Gateway in ipCurCfgGwTable. The range for ipCurCfgGwIndex is from 1 to ipGatewayTableMax. ipCurCfgGwAddr is the IP address of the default gateway.
altSwDefGwNotInService	Signifies that the default gateway is alive but not in service. ipCurCfgGwIndex is the index of the Gateway in ipCurCfgGwTable. The range for ipCurCfgGwIndex is from 1 to ipGatewayTableMax. ipCurCfgGwAddr is the IP address of the default gateway.
altSwVrrpNewMaster	Indicates that the sending agent has transitioned to "Master" state. vrrpCurCfgVirtRtrIndx is the VRRP virtual router table index referenced in vrrpCurCfgVirtRtrTable. The range is from 1 to vrrpVirtRtrTableMaxSize. vrrpCurCfgVirtRtrAddr is the VRRP virtual router IP address.
altSwVrrpNewBackup	Indicates that the sending agent has transitioned to "Backup" state. vrrpCurCfgVirtRtrIndx is the VRRP virtual router table index referenced in vrrpCurCfgVirtRtrTable. The range is from 1 to vrrpVirtRtrTableMaxSize. vrrpCurCfgVirtRtrAddr is the VRRP virtual router IP address.

Table 40. *Enterprise NOS-Supported Enterprise SNMP Traps (continued)*

Trap Name	Description
altSwVrrpAuthFailure	Signifies that a packet has been received from a router whose authentication key or authentication type conflicts with this router's authentication key or authentication type. Implementation of this trap is optional. vrrpCurCfgIfIndx is the VRRP interface index. This is equivalent to ifIndex in RFC 1213 mib. The range is from 1 to vrrpIfTableMaxSize. vrrpCurCfgIfPasswd is the password for authentication. It is a DisplayString of 0 to 7 characters.
altSwNtpNotServer	Signifies that the primary or secondary NTP server cannot be reached.
altSwNTPUpdateClock	Signifies that the system clock is updated with NTP server.
altSwECMPGatewayUp	Signifies that the ECMP gateway is up.
altSwECMPGatewayDown	Signifies that the ECMP gateway is down.
altSwOspfRouteUpdated	Signifies that an OSPF route update message was received.
altSwTempExceedThreshold	Signifies that the switch temperature has exceeded maximum safety limits.
altSwTempReturnThreshold	Signifies that the switch temperature has returned to under maximum safety limits.
altSwStackSwitchAttached	Signifies that a new switch has attached to the stack.
altSwStackSwitchDettached	Signifies that a new switch has detached from the stack.
altSwStackBackupPresent	Signifies that a new backup has been set.
altSwStackBackupGone	Signifies that the backup switch has been made unavailable.
altSwStackMasterAfterInit	Signifies that the switch has become master after init.

Table 40. *Enterprise NOS-Supported Enterprise SNMP Traps (continued)*

Trap Name	Description
altSwStackMasterFromBackup	Signifies that the switch has become master from backup.
altSwStackDuplicateJoinAttempt	Signifies that a new switch with duplicate UUID/bay has tried to join the stack.
altSwStackLinkUp	Signifies that a stack link has become up.
altSwStackLinkDown	Signifies that a stack link has become down.
altSwStackXferError	Signifies that a transfer between the master and a member has terminated with error.
altSwStackXferSuccess	Signifies that a transfer between the master and a member has terminated with no errors.
altSwStackSwitchTypeMismatch	Signifies that a new switch of different type has attempted to join the stack.
altSwStackImageSlotMismatch	Signifies that the slot of the boot image of a newly attached switch does not match that of the master.
altSwStackImageVersMismatch	Signifies that the version of the boot image of a newly attached switch does not match that of the master.
altSwStackBootCfgMismatch	Signifies that the booted config of a newly attached switch does not match that of the master.
altSwStackNvramMasterJoin	Signifies that a switch which was configured as a master in NVRAM has attached to the stack.
altSwStackForceDetach	Signifies that the master has sent a FORCE DETACH message to a member.
altVMGroupVMotion	Signifies that a virtual machine has moved from a port to another.
altVMGroupVMOnline	Signifies that an advance provisioned virtual machine has came online.

Table 40. *Enterprise NOS-Supported Enterprise SNMP Traps (continued)*

Trap Name	Description
altVMGroupVMVlanChange	Signifies that a virtual machine has entered a VLAN, or changed the VLAN.
vmCheckSpoofedvm	Signifies that a spoofed VM MAC was found.

Switch Images and Configuration Files

This section describes how to use MIB calls to work with switch images and configuration files. You can use a standard SNMP tool to perform the actions, using the MIBs listed in Table 41.

Table 41. lists the MIBS used to perform operations associated with the Switch Image and Configuration files.

Table 41. *MIBs for Switch Image and Configuration Files*

MIB Name	MIB OID
agTransferServer	1.3.6.1.4.1872.2.5.1.1.7.1.0
agTransferImage	1.3.6.1.4.1872.2.5.1.1.7.2.0
agTransferImageFileName	1.3.6.1.4.1872.2.5.1.1.7.3.0
agTransferCfgFileName	1.3.6.1.4.1872.2.5.1.1.7.4.0
agTransferDumpFileName	1.3.6.1.4.1872.2.5.1.1.7.5.0
agTransferAction	1.3.6.1.4.1872.2.5.1.1.7.6.0
agTransferLastActionStatus	1.3.6.1.4.1872.2.5.1.1.7.7.0
agTransferUserName	1.3.6.1.4.1872.2.5.1.1.7.9.0
agTransferPassword	1.3.6.1.4.1.1872.2.5.1.1.7.10.0
agTransferTSDumpFileName	1.3.6.1.4.1.1872.2.5.1.1.7.11.0

The following SNMP actions can be performed using the MIBs listed in Table 41.

- Load a new Switch image (boot or running) from a FTP/TFTP/SFTP server
- Load a previously saved switch configuration from a FTP/TFTP/SFTP server
- Save the switch configuration to a FTP/TFTP/SFTP server
- Save a switch dump to a FTP/TFTP/SFTP server

Loading a New Switch Image

To load a new switch image with the name "`MyNewImage-1.img`" into `image2`, follow the steps below. This example shows an FTP/TFTP/SFTP server at IPv4 address 192.168.10.10, though IPv6 is also supported.

1. Set the FTP/TFTP/SFTP server address where the switch image resides:

   ```
   Set agTransferServer.0 "192.168.10.10"
   ```

2. Set the area where the new image will be loaded:

   ```
   Set agTransferImage.0 "image2"
   ```

3. Set the name of the image:

   ```
   Set agTransferImageFileName.0 "MyNewImage-1.img"
   ```

4. If you are using an SFTP/FTP server, enter a username:

   ```
   Set agTransferUserName.0 "MyName"
   ```

5. If you are using an SFTP/FTP server, enter a password:

   ```
   Set agTransferPassword.0 "MyPassword"
   ```

6. Initiate the transfer. To transfer a switch image, enter 2 (gtimg):

   ```
   Set agTransferAction.0 "2"
   ```

Loading a Saved Switch Configuration

To load a saved switch configuration with the name "MyRunningConfig.cfg" into the switch, follow the steps below. This example shows a TFTP server at IPv4 address 192.168.10.10, though IPv6 is also supported.

1. Set the FTP/TFTP/SFTP server address where the switch Configuration File resides:

   ```
   Set agTransferServer.0 "192.168.10.10"
   ```

2. Set the name of the configuration file:

   ```
   Set agTransferCfgFileName.0 "MyRunningConfig.cfg"
   ```

3. If you are using an SFTP/FTP server, enter a username:

   ```
   Set agTransferUserName.0 "MyName"
   ```

4. If you are using an SFTP/FTP server, enter a password:

   ```
   Set agTransferPassword.0 "MyPassword"
   ```

5. Initiate the transfer. To restore a running configuration, enter 3:

   ```
   Set agTransferAction.0 "3"
   ```

Saving the Switch Configuration

To save the switch configuration to a FTP/TFTP/SFTP server follow the steps below. This example shows a FTP/TFTP/SFTP server at IPv4 address 192.168.10.10, though IPv6 is also supported.

1. Set the FTP/TFTP/SFTP server address where the configuration file is saved:

   ```
   Set agTransferServer.0 "192.168.10.10"
   ```

2. Set the name of the configuration file:

   ```
   Set agTransferCfgFileName.0 "MyRunningConfig.cfg"
   ```

3. If you are using an SFTP/FTP server, enter a username:

   ```
   Set agTransferUserName.0 "MyName"
   ```

4. If you are using an SFTP/FTP server, enter a password:

   ```
   Set agTransferPassword.0 "MyPassword"
   ```

5. Initiate the transfer. To save a running configuration file, enter 4:

   ```
   Set agTransferAction.0 "4"
   ```

Saving a Switch Dump

To save a switch dump to a FTP/TFTP/SFTP server, follow the steps below. This example shows an FTP/TFTP/SFTP server at 192.168.10.10, though IPv6 is also supported.

1. Set the FTP/TFTP/SFTP server address where the configuration will be saved:

   ```
   Set agTransferServer.0 "192.168.10.10"
   ```

2. Set the name of dump file:

   ```
   Set agTransferDumpFileName.0 "MyDumpFile.dmp"
   ```

3. If you are using an SFTP/FTP server, enter a username:

   ```
   Set agTransferUserName.0 "MyName"
   ```

4. If you are using an SFTP/FTP server, enter a password:

   ```
   Set agTransferPassword.0 "MyPassword"
   ```

5. Initiate the transfer. To save a dump file, enter 5:

   ```
   Set agTransferAction.0 "5"
   ```

Chapter 37. Service Location Protocol

Service Location Protocol (SLP) allows the switch to provide dynamic directory services that helps users find servers by attributes rather than by name or address. SLP eliminates the need for a user to know the name of a network host supporting a service. SLP allows the user to bind a service description to the network address of the service.

Service Location Protocol is described in RFC 2608.

Note: SLP is not supported on the internal management port (MGT).

SLP defines specialized components called agents that perform tasks and support services as follows:

- User Agent (UA) supports service query functions. It requests service information for user applications. The User Agent retrieves service information from the Service Agent or Directory Agents. A Host On-Demand client is an example of a User Agent.
- Service Agent (SA) provides service registration and service advertisement.

  Note: In this release, SA supports UA/DA on Linux with SLPv2 support.
- Directory Agent (DA) collects service information from Service Agents to provide a repository of service information in order to centralize it for efficient access by User Agents. There can only be one Directory Agent present per given host.

The Directory Agent acts as an intermediate tier in the SLP architecture, placed between the User Agents and the Service Agents, so they communicate only with the Directory Agent instead of with each other. This eliminates a large portion of the multicast request or reply traffic on the network, and it protects the Service Agents from being overwhelmed by too many service requests.

Services are described by the configuration of attributes associated with a type of service. A User Agent can select an appropriate service by specifying the attributes that it needs in a service request message. When service replies are returned, they contain a Uniform Resource Locator (URL) pointing to the service desired, and other information, such as server load, needed by the User Agent.

For more details on SLP configuration, see the *Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch Command Reference for Enterprise NOS 8.4*.

Active DA Discovery

When a Service Agent or User Agent initializes, it can perform Active Directory Agent Discovery using a multicast service request and specifies the special, reserved service type (`service:directory-agent`). Active DA Discovery is achieved through the same mechanism as any other discovery using SLP.

The Directory Agent replies with unicast service replies, which provides the URLs and attributes of the requested service.

Chapter 38. System License Keys

License keys determine the number of available ports on the CN4093. Each switch comes with basic license that provides the use of a limited number of physical ports. On top of the basic license, optional upgrade licenses can be installed to expand the number of available ports.

Obtaining Activation Keys

The upgrade licenses can be acquired using the *Lenovo System x Features on Demand* (FoD) website:

http://www.ibm.com/systems/x/fod/

You can also use the website to review and manage licenses, and to obtain additional help if required.

Note: An IBM ID and password are required to log into the FoD website. If you do not yet have an IBM ID, you can register at the website.

Activation keys are provided as files that must be uploaded to the CN4093. To acquire an activation key, use the FoD website to purchase an Authorization Code. You will need to provide the unique ID (UID) of the specific CN4093 where the key will be installed. The UID is the last 12 characters of the CN4093 serial number. This serial number is located on the Part Number (PN) label and is also displayed during successful login to the device.

When available, download the activation key file from the FoD site.

Installing Activation Keys

Once FoD activation key files have been acquired, they must be installed on the CN4093. The following example depicts use of the CN4093 Command Line Interface (CLI), but other device interfaces (such as SNMP) may also be used.

To install activation keys, complete the following steps:

a. Log in to the CN4093.

b. At the CLI prompt, enter the following commands:

```
CN4093> enable
CN4093# configure terminal
CN4093(config)# software-key
CN4093(config)# enakey addr <server IP address> keyfile <key filename>
```

c. Follow the prompts to enter the appropriate parameters, including the file transfer protocol and server parameters.

Note: Repeat the **enakey** command for any additional keys being installed.

The system prompts you to confirm your request.

Transferring Activation Keys

Licenses keys are based on the unique CN4093 device serial number and are non-transferable.

In the event that the CN4093 must be replaced, a new activation key must be acquired and installed. When the replacement is handled through Lenovo Service and Support, your original license will be transferred to the serial number of the replacement unit and you will be provided a new license key.

Trial Keys

Trial keys are license keys used for evaluation purposes, upgrading the number of available ports for limited time. They are managed and obtained like regular license keys, from the *Lenovo System x Features on Demand* (FoD) website:

http://www.ibm.com/systems/x/fod/

Trial keys expire after a predefined number of days. 10 days before the expiration date, the switch will begin to issue the following syslog messages:

```
The software demo license for Upgrade1 will expire in 10 days. The switch
will automatically set the default port-map after the license expires.
This will cause ports that are not in the default port-map to lose their
configuration. Please backup your configuration or enter a valid license
key so the configuration will not be altered.
```

When the trial license expires, all features enabled by the key are disabled, configuration files (active and backup) are deleted and the switch revert to the default port-map. To prevent this, either install a regular upgrade license to overwrite the trial key or manually remove the trial key.

Once a trial key is installed, it cannot be reused.

Flexible Port Mapping

Flexible Port Mapping allows administrators to manually enable or disable specific switch ports within the limitations of the installed licenses' bandwidth.

For instance, the FlexSystem may include two compute nodes and a single QSFP+ uplink, while the current license has the INTA1 – INTA14 and EXT1 – EXT10 Ethernet ports enabled by default.

To make best use of the available resources, the administrator decides to activate internal ports INTB1, INTB2, INTC1 and INTC2 to provide redundant connections for the two compute nodes and to enable the high speed QSFP+ EXT3 port for the uplink.

The total bandwidth required for this operation amounts to 80 Gbps (40 Gbps for the four additional 10 Gbps internal ports and 40 Gbps for the additional external QSFP+ port). The administrator decides to allocate this bandwidth by deactivating 6 internal and 2 external 10 Gbps ports.

To implement the above scenario, follow these steps:

a. Deactivate the ports required to clear the 80 Gbps required bandwidth:

```
CN4093(config)# no boot port-map INTA9
CN4093(config)# no boot port-map INTA10
CN4093(config)# no boot port-map INTA11
CN4093(config)# no boot port-map INTA12
CN4093(config)# no boot port-map INTA13
CN4093(config)# no boot port-map INTA14
CN4093(config)# no boot port-map EXT9
CN4093(config)# no boot port-map EXT10
```

b. Activate the required ports:

```
CN4093(config)# boot port-map INTB1
CN4093(config)# boot port-map INTC1
CN4093(config)# boot port-map INTB2
CN4093(config)# boot port-map INTC2
CN4093(config)# boot port-map EXT3
```

c. To verify the configuration, run the following command:

```
CN4093(config)# show boot port-map
```

Flexible Port Mapping is disabled if all available licenses are installed (all physical ports are available).

Removing a license key reverts the port mapping to the default settings for the remaining licensing level. To manually revert the port mapping to the default settings use the following command:

```
CN4093(config)# default boot port-map
```

Chapter 39. Secure Input/Output Module

The Secure Input/Output Module (SIOM) enables you to determine which protocols can be enabled. The SIOM only allows secured traffic and secured authentication management.

The following topics are discussed in this chapter:

SIOM Overview

In networking solutions, a new approach about adopting a security level on Input/Output modules has been developed. This security level encompasses secured authentication management and only allows secure traffic and protocols.

IOMs can be classified into two security categories:

- Legacy Input/Output Modules (LIOMs)

  LIOMs are not capable of provisioning any security policy setting. All IOMs developed before the SIOM feature was introduced are of type LIOM.

- Secure Input/Output Modules (SIOMs)

  SIOMs have security characteristics that allow them to integrate the network assigned security policy.

For IOM to be in SIOM mode, both the IOM and the CMM (Chassis Management Module) containing it must be running SIOM-capable software, and the IOM must have SIOM enabled. In all other cases, the IOM operates in LIOM mode.

When the IOM is in SIOM mode, the security characteristics configured on the CMM are sent to the IOM. These characteristics can be divided into the following categories:

- Policy setting
- User Account Management
- Secure LDAP (LDAPS) authentication

To see whether SIOM is enabled on the IOM, use the following command:

```
CN 4093(config)# show boot siom
Current SIOM setting: disabled
Saved SIOM setting:   disabled
```

This shows both the current SIOM setting and the saved setting that will be applied after reboot.

SIOM is disabled by default. To enable SIOM on the switch, in Global Configuration mode, enter:

```
CN 4093(config)# boot siom enable
```

To disable SIOM, enter:

```
CN 4093(config)# no boot siom enable
```

Note: You must reboot the switch for SIOM settings to take effect.

Switch Access in SIOM Mode

After the embedded switch is provisioned by the CMM in the SIOM mode, the switch will automatically update its LDAP settings (`startTLS`, `LDAPS` or `LDAP`) to the ones configured on the CMM. When no external LDAP server is configured on the CMM, CMM itself will serve as the local LDAP server. The LDAP client configured on the CMM is pushed onto the switch and the LDAP credentials used to access the CMM can also be used to access the switch.

To access the switch, you may now use one of the following methods:

- The CMM credentials
- Other user credentials which depend on the SIOM security policy setting, as follows:
  - In legacy mode, if RADIUS or TACACS+ is enabled, they will replace LDAP as the authentication method. If LDAP backdoor mode is enabled, you can still use local authentication by using `noldap` as the username.
  - In secure mode, you may use the provisioned LDAP credentials.

Notes:

- Once the switch is provisioned by the CMM in SIOM mode, it cannot be accessed using the switch local user accounts.
- The switch may perform an additional reboot automatically after changing the SIOM state or upgrading the CMM software.

Using SIOM with Stacking

In stacking mode, configuring SIOM is only supported on the Master switch. Hence, the command:

```
CN 4093# [no] boot siom enable
```

is only supported on the Master switch. On stack member switches, SIOM is configured by the Master switch, and the member switches automatically inherit the Master switch SIOM setting. When upgrading to SIOM-capable software:

- The Master, Backup, and member switches need to be rebooted for SIOM to take effect.
- When SIOM is enabled on the Master, it is applied on all stack members automatically.
- If a new switch with a different boot SIOM configuration is attached to the stack, the switch will inherit the boot SIOM configuration from the Master and will automatically reboot.
- When two stacks are joined, the selected Master for the two stacks will push its own boot SIOM configuration, and the added members will automatically reboot.
- There will be no changes in the SIOM policy on members if the stack is split.

Note: Lenovo recommends using staggered upgrade. In this case, the upgrade will take more time, depending on how large the stack setup is, but the traffic loss will be minimal.

Overall, in a stacking mode, it takes longer (about 9 minutes) to reload SIOM-enabled software than a non-SIOM enabled software. If staggered-upgrade procedure is used, this duration increases according to the number of switches in the stack.

The process of upgrading from non-SIOM enabled software to SIOM-enabled software takes about 15 minutes. If a staggered-upgrade procedure is used, this duration increases according to the number of switches in the stack.

If the Master switch gets rebooted, the Backup switch becomes the Master (operation called Master failover) and it will be SIOM provisioned. If the SIOM provisioning occurs for the first time on this switch, it will also reboot for the LIOM to SIOM transition. The new Master will reboot only after the Backup switch (original Master) rejoins the stack.

SIOM Feature Considerations

SIOM has two aspects which must be accomplished on the switch:

- The provisioning process which supplies all the necessary settings for the secure Ethernet connection for management and for the secure protocols enabled on the switch.
- The protocols enabled during the functioning of a switch in SIOM.

Switch boots up with all operational (data) ports disabled. Although the management ports are enabled, they can't be used by `admin` to set up the switch until the configuration is applied.

Internal management port is used by the CMM during the provisioning to exchange information with IOM.

At the end of provisioning, when SIOM is enabled, the rest of the operational ports come up and the switch will be fully functional.

When in SIOM mode, the PKI system of switch cannot be controlled. The user cannot import his own certificate. All certificates are provisioned by CMM.

Creating a Policy Setting

The policy setting can be either secure (IOM is in secure mode) or legacy (IOM is in legacy mode). In secure mode, only communication protocols that are deemed secure can be used; most protocols that are not deemed secure are disabled. In legacy mode setting, all protocols are allowed (LIOM behavior).

To display the current policy setting, enter:

```
CN 4093(config)# show boot security-policy
```

Note: Security policy can be applied only from CMM. You must reboot the IOM for a new policy setting to be applied.

Protocols Affected by the Policy Setting

This section explains which protocols can and cannot operate in secure mode on the CN4093 10 Gb Converged Scalable Switch.

Insecure Protocols

When you are in Secure Mode, the following protocols are deemed "insecure" and are disabled:

- HTTP
- LDAP Client
- SNMPv1
- SNMPv2
- Telnet (server and client)
- FTP (server and client)
- Radius (client)
- TFTP Server

Except for the TFTP server, these protocols cannot be enabled when the switch is operating in Secure Mode because the commands to enable or disable them are no longer enabled.

The following protocols, although deemed "insecure," are enabled by default and can be disabled.

- DHCP client
- SysLog

Note: Service Location Protocol (SLP) Discovery is also deemed "insecure" but is unaffected by Secure Mode. SLP has the same default settings as in Legacy Mode. If you can enable or disable SLP in Legacy Mode, you can enable or disable it the same way in Secure Mode.

The following supported protocols are not enabled by default but can always be enabled in Secure Mode.

- DNS Resolution

- TFTP client (for signed items only, such as switch images)

Secure Protocols

The following protocols are deemed "secure" and are enabled by default in Secure Mode:

- SCP Server
- SNMPv3 Client
- SFTP Client
- SSHv2 Server
- SSHv2 Client
- HTTPS Server
- TACACS+ Client

You can disable these protocols.

The following protocols are deemed "secure" and cannot be disabled in any mode:

- NTP Client v4
- LDAPS Client

The following protocols are also deemed "secure" on the CN4093 and can be enabled.

- IKE
- IPSec

The default state for these protocols in Secure Mode, whether enabled or disabled, is the same as in Legacy Mode.

The following protocols are deemed "secure" but are not currently supported by the CN4093:

- EAPoL
- SCP
- S/MIME
- SNMPv3 Manager
- TCP command secure mode (Port 6091)

Insecure Protocols Unaffected by SIOM

The following protocols are deemed "insecure" but can be enabled in all Security Policy Modes:

- Ping
- Ping IPv6
- Traceroute
- Traceroute IPv6
- TFTP IPv6

- SNMPv3 IPv6
- bootp

Notes:

- Telnet IPv6 and TFTP IPv6 are disabled in Secure Mode.
- TFTP IPv6 is allowed in Secure Mode for signed image transfers only.

Managing User Accounts

SNMPv3 user accounts with customized attributes can be created on the CMM and pushed to the IOM. For each SNMPv3 user account created on the CMM, the IOM creates a local SNMPv3 user account. The SNMPv3 user database then creates new user-per-profile user lists. It then uses this database to authenticate users.

Note: SNMPv3 does not support LDAP user management, so the CMM must provision SNMPv3 user accounts to the IOM.

Using Centralized SNMPv3 Management with SIOM

There is a setting on the CMM to indicate whether the SNMPv3 centralized user management is enabled; this is called the *Centralized Flag*.

When the IOM runs as SIOM and the Centralized Flag is enabled, SNMPv3 will enable Node Accounts and will disable Local Accounts. When the IOM runs as LIOM *or* the Centralized Flag is disabled, SNMPv3 will use Local Accounts and disable Node Accounts. Node Accounts represent accounts configured on the CMM, while Local Accounts are accounts configured on the IOM.

Since there is no case where both the Node Account and Local Account are enabled, the username of a Node Account can be duplicated with a Local Account username.

Implementing SNMPv3 with SIOM

The following commands are available for implementing SNMPv3 with SIOM:

- `access snmp read-only`
- `access snmp read-write`
- `snmp-server access`
- `snmp-server community`
- `snmp-server group`
- `snmp-server host`
- `snmp-server notify`
- `snmp-server read-community`
- `snmp-server read-community-additional`
- `snmp-server target-address`
- `snmp-server target-parameters`
- `snmp-server user`
- `snmp-server version`
- `snmp-server view`
- `snmp-server write-community`
- `snmp-server write-community-additional`
- `show snmp-server v3`

For more information about these commands, see the *Lenovo ISCLI—Industry Standard CLI Command Reference for the Lenovo Flex System Fabric CN4093 10 Gb Converged Scalable Switch*.

Implementing Secure LDAP (LDAPS)

Lightweight Directory Access Protocol (LDAP) is a protocol for accessing distributed directory information services over a network. Enterprise NOS uses LDAP for authentication and authorization. With an LDAP client enabled, the switch will authenticate a user and determine the user's privilege level by checking with one or more directory servers instead of a local database of users. This prevents customers from having to configure local user accounts on multiple switches; they can maintain a centralized directory instead.

As part of SIOM, you can implement Secure Lightweight Directory Access Protocol (LDAPS) in addition to standard LDAP.

Enabling LDAPS

When the IOM is in SIOM mode, all LDAP configurations are made from the CMM and pushed to the IOM. When the IOM is in LIOM mode, the CLI can be used to configure LDAP settings. LDAPS is disabled by default. To enable LDAPS:

1. Turn LDAP authentication on

```
CN 4093(config)# ldap-server enable
```

2. Enable LDAP Enhanced Mode:

```
CN 4093(config)# ldap-server mode enhanced
```

This changes the `ldap-server` subcommands to support LDAPS.

3. Configure the IPv4 addresses of each LDAP server.

```
CN 4093(config)# ldap-server host {1-4} <IP address or hostname>
```

4. You may change the default TCP port number used to listen to LDAPS (optional).

The well-known port for LDAP is 636.

```
CN 4093(config)# ldap-server port <1-65000>
```

5. Configure the Security Mode:

```
CN 4093(config)# ldap-server security {clear|ldaps|mutual|starttls}
```

where:

Parameter	Description
clear	Cleartext Mode (no security)
ldaps	LDAPS Mode
mutual	Mutual authentication in Transport Layer Security (TLS)
starttls	Secure LDAP via StartTLS without cleartext fallback

6. Configure the distinguished name (DN) and password (optional).

```
CN 4093(config)# ldap-server binddn dn "<distinguished name> "
CN 4093(config)# ldap-server binddn key "<password> "
```

If this is not configured, the switch will use user-provided login credentials to bind. A DN will then be constructed from the user's login credentials and then used in the initial BIND attempt.

7. Configure the root DN:

```
CN 4093(config)# ldap-server basedn <root DN name>
```

8. Configure the user search attribute (optional):

```
CN 4093(config)# ldap-server attribute username <search attribute>
```

If no user search attribute is specified, the default is `uid`.

9. Configure the group search attribute (optional):

```
CN 4093(config)# ldap-server attribute group <search attribute>
```

If no group search attribute is specified, the default is `memberOf`.

10. Configure the login permissions attribute:

```
CN 4093(config)# ldap-server attribute login-permission <attribute>
```

Note: If no login permissions attribute is configured, LDAP client will not function.

11. Configure the group filter attribute (optional):

```
CN 4093(config)# ldap-server group-filter <filter attributes separated by comma>
```

Note: The group filter string must contain no whitespace.

If no group filter attribute is configured, no groups will be filtered and all groups will be considered in any search.

12. Enable DNS server verification:

```
CN 4093(config)# ldap-server srv
```

Disabling LDAPS

To disable LDAPS, enter:

```
CN 4093(config)# ldap-server security clear
CN 4093(config)# ldap-server mode legacy
```

For information about using LDAP in Legacy Mode, see "LDAP Authentication and Authorization" on page 114.

Syslogs and LDAPS

Syslogs are displayed for the following error conditions:

- `Password change required on first login`
- `Password expired`
- `Username or password invalid`
- `Account temporarily locked`
- `Unknown/no reason given`

SIOM Dependencies

The following points are relevant to SIOM:

- The CMM has a Certificate Authority (CA) capable of signing the certificates involved for authenticating the IOM in the SSL and TLS processes and protocols.
- The correctness of the configuration depends upon the settings on the CMM. This is especially important for NTP and LDAP, which ensure switch operability. For example, if the LDAP client is configured incorrectly, the switch cannot be managed.
- The Enhanced Configuration and Management (EHCM) module configures the NTP client. Therefore, the NTP client is dependent upon the ECHM module being enabled and functional.
- Some protocols cannot be changed from enabled to disabled without restarting the switch. The IOM may reboot when switching between the SIOM and LIOM.

Part 8: Monitoring

The ability to monitor traffic passing through the CN4093 can be invaluable for troubleshooting some types of networking problems. This sections cover the following monitoring features:

- Remote Monitoring (RMON)
- sFLOW
- Port Mirroring

Chapter 40. Remote Monitoring

Remote Monitoring (RMON) allows network devices to exchange network monitoring data.

RMON performs the following major functions:

- Gathers cumulative statistics for Ethernet interfaces
- Tracks a history of statistics for Ethernet interfaces
- Creates and triggers alarms for user-defined events

RMON Overview

The RMON MIB provides an interface between the RMON agent on the switch and an RMON management application. The RMON MIB is described in RFC 1757.

The RMON standard defines objects that are suitable for the management of Ethernet networks. The RMON agent continuously collects statistics and proactively monitors switch performance. RMON allows you to monitor traffic flowing through the switch.

The switch supports the following RMON Groups, as described in RFC 1757:

- RMON Group 1–Statistics
- RMON Group 2–History
- RMON Group 3–Alarms
- RMON Group 9–Events

RMON Group 1–Statistics

The switch supports collection of Ethernet statistics as outlined in the RMON statistics MIB, in reference to `etherStatsTable`. RMON statistics are sampled every second, and new data overwrites any old data on a given port.

Note: RMON port statistics must be enabled for the port before you can view RMON statistics.

To configure RMON Statistics:

1. Enable RMON on each port where you wish to collect RMON statistics.

```
CN 4093(config)# interface port 23
CN 4093(config-if)# rmon
```

2. View RMON statistics for the port.

```
CN 4093(config-if)# show interface port 23 rmon-counters
----------------------------------------------------------------------
RMON statistics for port 23:
etherStatsDropEvents:                       NA
etherStatsOctets:                      7305626
etherStatsPkts:                          48686
etherStatsBroadcastPkts:                  4380
etherStatsMulticastPkts:                  6612
etherStatsCRCAlignErrors:                   22
etherStatsUndersizePkts:                     0
etherStatsOversizePkts:                      0
etherStatsFragments:                         2
etherStatsJabbers:                           0
etherStatsCollisions:                        0
etherStatsPkts64Octets:                  27445
etherStatsPkts65to127Octets:             12253
etherStatsPkts128to255Octets:             1046
etherStatsPkts256to511Octets:              619
etherStatsPkts512to1023Octets:            7283
etherStatsPkts1024to1518Octets:             38
```

RMON Group 2–History

The RMON History Group allows you to sample and archive Ethernet statistics for a specific interface during a specific time interval.

Note: RMON port statistics must be enabled for the port before an RMON history group can monitor the port.

Data is stored in buckets, which store data gathered during discreet sampling intervals. At each configured interval, the history instance takes a sample of the current Ethernet statistics, and places them into a bucket. History data buckets reside in dynamic memory. When the switch is re-booted, the buckets are emptied.

Requested buckets are the number of buckets, or data slots, requested by the user for each History Group. Granted buckets are the number of buckets granted by the system, based on the amount of system memory available. The system grants a maximum of 50 buckets.

Use an SNMP browser to view History samples.

History MIB Objects

The type of data that can be sampled must be of an `ifIndex` object type, as described in RFC1213 and RFC1573. The most common data type for the history sample is as follows:

`1.3.6.1.2.1.2.2.1.1.`<*x*>
`-mgmt.interfaces.ifTable.ifIndex.interface`

The last digit (*x*) represents the interface on which to monitor, which corresponds to the switch port number. History sampling is done per port, by utilizing the interface number to specify the port number.

Configuring RMON History

This example configuration creates an RMON History Group to monitor port 1. It takes a data sample every two minutes, and places the data into one of the 30 requested buckets. After 30 samples are gathered, the new samples overwrite the previous samples, beginning with the first bucket.

1. Enable RMON on each port where you wish to collect RMON History.

```
CN 4093(config)# interface port 1
CN 4093(config-if)# rmon
CN 4093(config-if)# exit
```

2. Configure the RMON History parameters.

```
CN 4093(config)# rmon history 1 interface-oid 1.3.6.1.2.1.2.2.1.1.<x>
CN 4093(config)# rmon history 1 requested-buckets 30
CN 4093(config)# rmon history 1 polling-interval 120
CN 4093(config)# rmon history 1 owner "rmon port 1 history"
```

where <x> is the number of the port to monitor. For example, the full OID for port 1 would be: `1.3.6.1.2.1.2.2.1.1.1`

3. View RMON history for the port.

```
CN 4093(config)# show rmon history
RMON History group configuration:

Index          IFOID            Interval   Rbnum  Gbnum
-----  ----------------------  --------   -----  -----
    1  1.3.6.1.2.1.2.2.1.1.1         120      30     30

Index                        Owner
-----  -------------------------------------------
    1  rmon port 1 history
```

RMON Group 3–Alarms

The RMON Alarm Group allows you to define a set of thresholds used to determine network performance. When a configured threshold is crossed, an alarm is generated. For example, you can configure the switch to issue an alarm if more than 1,000 CRC errors occur during a 10-minute time interval.

Each Alarm index consists of a variable to monitor, a sampling time interval, and parameters for rising and falling thresholds. The Alarm group can be used to track rising or falling values for a MIB object. The object must be a counter, gauge, integer, or time interval.

Use one of the following commands to correlate an Alarm index to an Event index:

```
CN 4093(config)# rmon alarm <alarm number> rising-crossing-index <event number>
CN 4093(config)# rmon alarm <alarm number> falling-crossing-index
<event number>
```

Alarm MIB Objects

The most common data types used for alarm monitoring are `ifStats`: errors, drops, bad CRCs, and so on. These MIB Object Identifiers (OIDs) correlate to the ones tracked by the History group. An example of an ICMP stat is as follows:

`1.3.6.1.2.1.5.1.`*<x>* `- mgmt.icmp.icmpInMsgs`

where *x* represents the interface on which to monitor, which corresponds to the switch interface number or port number, as follows:

- 1 through 128 = Switch interface number
- 129 = Switch port 1
- 130 = Switch port 2
- 131 = Switch port 3, and so on.

This value represents the alarm's MIB OID, as a string. Note that for non-tables, you must supply a `.0` to specify an end node.

Configuring RMON Alarms

Alarm Example 1

This example configuration creates an RMON alarm that checks `ifInOctets` on port 20 once every hour. If the statistic exceeds two billion, an alarm is generated that triggers event index 6.

Configure the RMON Alarm parameters to track the number of packets received on a port.

```
CN 4093(config)# rmon alarm 1 oid 1.3.6.1.2.1.2.2.1.10.129
CN 4093(config)# rmon alarm 1 alarm-type rising
CN 4093(config)# rmon alarm 1 rising-crossing-index 100
CN 4093(config)# rmon alarm 1 interval 3600
CN 4093(config)# rmon alarm 1 rising-limit 2000000000
CN 4093(config)# rmon alarm 1 owner "Alarm for ifInOctets"
```

Alarm Example 2

This example configuration creates an RMON alarm that checks `icmpInEchos` on the switch once every minute. If the statistic exceeds 200 within a 60 second interval, an alarm is generated that triggers event index 5.

Configure the RMON Alarm parameters to track ICMP messages.

```
CN 4093(config)# rmon alarm 1 oid 1.3.6.1.2.1.5.8.0
CN 4093(config)# rmon alarm 1 alarm-type rising
CN 4093(config)# rmon alarm 1 rising-crossing-index 110
CN 4093(config)# rmon alarm 1 interval-time 60
CN 4093(config)# rmon alarm 1 rising-limit 200
CN 4093(config)# rmon alarm 1 sample delta
CN 4093(config)# rmon alarm 1 owner "Alarm for icmpInEchos"
```

RMON Group 9–Events

The RMON Event Group allows you to define events that are triggered by alarms. An event can be a log message, an SNMP trap message, or both.

When an alarm is generated, it triggers a corresponding event notification. Use the following commands to correlate an Event index to an alarm:

```
CN 4093(config)# rmon alarm <alarm number> rising-crossing-index <event number>
CN 4093(config)# rmon alarm <alarm number> falling-crossing-index
<event number>
```

RMON events use SNMP and system logs to send notifications. Therefore, an SNMP trap host must be configured for trap event notification to work properly.

RMON uses a syslog host to send syslog messages. Therefore, an existing syslog host must be configured for event log notification to work properly. Each log event generates a system log message of type RMON that corresponds to the event.

For example, to configure the RMON event parameters.

```
CN 4093(config)# rmon event 110 type log
CN 4093(config)# rmon event 110 description "SYSLOG_this_alarm"
CN 4093(config)# rmon event 110 owner "log icmpInEchos alarm"
```

This configuration creates an RMON event that sends a syslog message each time it is triggered by an alarm.

Chapter 41. sFLOW

The CN4093 supports sFlow technology for monitoring traffic in data networks. The switch includes an embedded sFlow agent which can be configured to sample network traffic and provide continuous monitoring information of IPv4 traffic to a central sFlow analyzer.

The switch is responsible only for forwarding sFlow information. A separate sFlow analyzer is required elsewhere on the network in order to interpret sFlow data.

Note: Enterprise NOS 8.4 does not support IPv6 for sFLOW.

sFlow Statistical Counters

The CN4093 can be configured to send network statistics to an sFlow analyzer at regular intervals. For each port, a polling interval of 5 to 60 seconds can be configured, or 0 (the default) to disable this feature.

When polling is enabled, at the end of each configured polling interval, the CN4093 reports general port statistics and port Ethernet statistics.

sFlow Network Sampling

In addition to statistical counters, the CN4093 can be configured to collect periodic samples of the traffic data received on each port. For each sample, 128 bytes are copied, UDP-encapsulated, and sent to the configured sFlow analyzer.

For each port, the sFlow sampling rate can be configured to occur once each 256 to 65536 packets, or 0 to disable (the default). A sampling rate of 256 means that one sample will be taken for approximately every 256 packets received on the port. The sampling rate is statistical, however. It is possible to have slightly more or fewer samples sent to the analyzer for any specific group of packets (especially under low traffic conditions). The actual sample rate becomes most accurate over time, and under higher traffic flow.

sFlow sampling has the following restrictions:

- Sample Rate—The fastest sFlow sample rate is 1 out of every 256 packets.
- ACLs—sFlow sampling is performed before ACLs are processed. For ports configured both with sFlow sampling and one or more ACLs, sampling will occur regardless of the action of the ACL.
- Port Mirroring—sFlow sampling will not occur on mirrored traffic. If sFlow sampling is enabled on a port that is configured as a port monitor, the mirrored traffic will not be sampled.

Note: Although sFlow sampling is not generally a CPU-intensive operation, configuring fast sampling rates (such as once every 256 packets) on ports under heavy traffic loads can cause switch CPU utilization to reach maximum. Use larger rate values for ports that experience heavy traffic.

sFlow Example Configuration

1. Specify the location of the sFlow analyzer (the server and optional port to which the sFlow information will be sent):

```
CN 4093(config)# sflow server <IPv4 address>(sFlow server address)
CN 4093(config)# sflow port <service port> (Set the optional service port)
CN 4093(config)# sflow enable              (Enable sFlow features)
```

   By default, the switch uses established sFlow service port 6343.

   To disable sFlow features across all ports, use the following command:
   `CN 4093(config)# no sflow enable`

2. On a per-port basis, define the statistics polling rate:

```
CN 4093(config)# interface port <port>
CN 4093(config-if)# sflow polling <polling rate>(Statistics polling rate)
```

   Specify a polling rate between 5 and 60 seconds, or 0 to disable. By default, polling is 0 (disabled) for each port.

3. On a per-port basis, define the data sampling rate:

```
CN 4093(config-if)# sflow sampling <sampling rate>(Data sampling rate)
```

   Specify a sampling rate between 256 and 65536 packets, or 0 to disable. By default, the sampling rate is 0 (disabled) for each port.

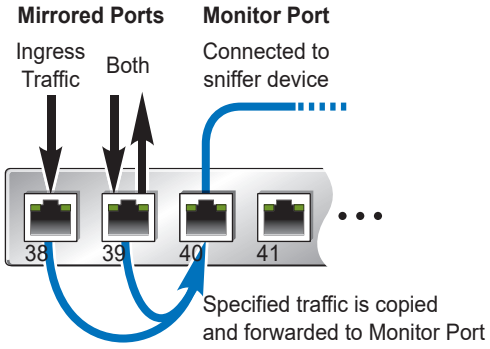
4. Save the configuration.

Chapter 42. Port Mirroring

The Enterprise NOS port mirroring feature allows you to mirror (copy) the packets of a target port, and forward them to a monitoring port. Port mirroring functions for all layer 2 and layer 3 traffic on a port. This feature can be used as a troubleshooting tool or to enhance the security of your network. For example, an IDS server or other traffic sniffer device or analyzer can be connected to the monitoring port in order to detect intruders attacking the network.

The CN4093 supports a "many to one" mirroring model. As shown in Figure 65, selected traffic for ports EXT1 and EXT2 is being monitored by port EXT3. In the example, both ingress traffic and egress traffic on port EXT2 are copied and forwarded to the monitor. However, port EXT1 mirroring is configured so that only ingress traffic is copied and forwarded to the monitor. A device attached to port EXT3 can analyze the resulting mirrored traffic.

Figure 65. Mirroring Ports



In standalone (non-stacking) mode, the CN4093 supports two monitor ports with two-way mirroring, or four monitor ports with one-way mirroring. In stacking mode, one monitor port with two-way mirroring, or two monitor ports with one-way mirroring is supported. Each monitor port can receive mirrored traffic from any number of target ports.

Enterprise NOS does not support "one to many" or "many to many" mirroring models where traffic from a specific port traffic is copied to multiple monitor ports. For example, port EXT1 traffic cannot be monitored by both port EXT3 and EXT4 at the same time, nor can port EXT2 ingress traffic be monitored by a different port than its egress traffic.

Ingress and egress traffic is duplicated and sent to the monitor port after processing.

Note: The CN4093 10 Gb Converged Scalable Switch (CN4093) cannot mirror LACPDU packets. Also, traffic on management VLANs is not mirrored to the external ports.

Port Mirroring Behavior

This section describes the composition of monitored packets in the CN4093, based on the configuration of the ports.

- Packets mirrored at port egress are mirrored prior to VLAN tag processing and may have a different PVID than packets that egress the port toward their actual network destination.
- Packets mirrored at port ingress are not modified.

Configuring Port Mirroring

The following procedure may be used to configure port mirroring for the example shown in Figure 65 on page 611:

1. Specify the monitoring port, the mirroring port(s), and the port-mirror direction.

```
CN 4093(config)# port-mirroring monitor-port EXT3 mirroring-port EXT1 in
CN 4093(config)# port-mirroring monitor-port EXT3 mirroring-port EXT2 both
```

2. Enable port mirroring.

```
CN 4093(config)# port-mirroring enable
```

3. View the current configuration.

```
CN 4093# show port-mirroring            (Display the current settings)
Port mirroring is enabled
Monitoring Ports    Mirrored Ports
INTA1                 none
INTA2                 none
INTA3                 none
INTA4                 none
...
EXT1                  none
EXT2                  none
EXT3                  EXT1, in
                    EXT2, both
EXT4                  none
...
```

Part 9: Appendices

Appendix A. Glossary

DIP	The destination IP address of a frame.
Dport	The destination port (application socket: for example, http-80/https-443/DNS-53)
HBA	Host Bus Adapter. An adapter or card that interfaces with device drivers in the host operating system and the storage target in a Storage Area Network (SAN). It is equivalent to a Network Interface Controller (NIC) from a Local Area Network (LAN).
NAT	Network Address Translation. Any time an IP address is changed from one source IP or destination IP address to another address, network address translation can be said to have taken place. In general, half NAT is when the destination IP or source IP address is changed from one address to another. Full NAT is when both addresses are changed from one address to another. No NAT is when neither source nor destination IP addresses are translated.
Preemption	In VRRP, preemption will cause a Virtual Router that has a lower priority to go into backup should a peer Virtual Router start advertising with a higher priority.
Priority	In VRRP, the value given to a Virtual Router to determine its ranking with its peer(s). Minimum value is 1 and maximum value is 254. Default is 100. A higher number will win out for master designation.
Proto (Protocol)	The protocol of a frame. Can be any value represented by a 8-bit value in the IP header adherent to the IP specification (for example, TCP, UDP, OSPF, ICMP, and so on.)
SIP	The source IP address of a frame.
SPort	The source port (application socket: for example, HTTP-80/HTTPS-443/DNS-53).
Tracking	In VRRP, a method to increase the priority of a virtual router and thus master designation (with preemption enabled). Tracking can be very valuable in an active/active configuration. You can track the following: • Active IP interfaces on the Web switch (increments priority by 2 for each) • Active ports on the same VLAN (increments priority by 2 for each) • Number of virtual routers in master mode on the switch
VIR	Virtual Interface Router. A VRRP address is an IP interface address shared between two or more virtual routers.
Virtual Router	A shared address between two devices utilizing VRRP, as defined in RFC 2338. One virtual router is associated with an IP interface. This is one of the IP interfaces that the switch is assigned. All IP interfaces on the CN4093s must be in a VLAN. If there is more than one VLAN defined on the Web switch, then the VRRP broadcasts will only be sent out on the VLAN of which the associated IP interface is a member.

VRID	Virtual Router Identifier. In VRRP, a numeric ID is used by each virtual router to create its MAC address and identify its peer for which it is sharing this VRRP address. The VRRP MAC address as defined in the RFC is 00-00-5E-00-01-<*VRID*>. If you have a VRRP address that two switches are sharing, then the VRID number needs to be identical on both switches so each virtual router on each switch knows with whom to share.
VRRP	Virtual Router Redundancy Protocol. A protocol that acts very similarly to Cisco's proprietary HSRP address sharing protocol. The reason for both of these protocols is so devices have a next hop or default gateway that is always available. Two or more devices sharing an IP interface are either advertising or listening for advertisements. These advertisements are sent via a broadcast message to an address such as 224.0.0.18. With VRRP, one switch is considered the master and the other the backup. The master is always advertising via the broadcasts. The backup switch is always listening for the broadcasts. Should the master stop advertising, the backup will take over ownership of the VRRP IP and MAC addresses as defined by the specification. The switch announces this change in ownership to the devices around it by way of a Gratuitous ARP, and advertisements. If the backup switch didn't do the Gratuitous ARP the Layer 2 devices attached to the switch would not know that the MAC address had moved in the network. For a more detailed description, refer to RFC 2338.

Appendix B. Getting help and technical assistance

If you need help, service, or technical assistance or just want more information about Lenovo products, you will find a wide variety of sources available from Lenovo to assist you.

Use this information to obtain additional information about Lenovo and Lenovo products, and determine what to do if you experience a problem with your Lenovo system or optional device.

Note: This section includes references to IBM web sites and information about obtaining service. IBM is Lenovo's preferred service provider for the System x, Flex System, and NeXtScale System products.

Before you call, make sure that you have taken these steps to try to solve the problem yourself.

If you believe that you require warranty service for your Lenovo product, the service technicians will be able to assist you more efficiently if you prepare before you call.

- Check all cables to make sure that they are connected.
- Check the power switches to make sure that the system and any optional devices are turned on.
- Check for updated software, firmware, and operating-system device drivers for your Lenovo product. The Lenovo Warranty terms and conditions state that you, the owner of the Lenovo product, are responsible for maintaining and updating all software and firmware for the product (unless it is covered by an additional maintenance contract). Your service technician will request that you upgrade your software and firmware if the problem has a documented solution within a software upgrade.
- If you have installed new hardware or software in your environment, check the IBM ServerProven website to make sure that the hardware and software is supported by your product.
- Go to the IBM Support portal to check for information to help you solve the problem.
- Gather the following information to provide to the service technician. This data will help the service technician quickly provide a solution to your problem and ensure that you receive the level of service for which you might have contracted.
  - Hardware and Software Maintenance agreement contract numbers, if applicable
  - Machine type number (Lenovo 4-digit machine identifier)
  - Model number
  - Serial number
  - Current system UEFI and firmware levels
  - Other pertinent information such as error messages and logs

- Start the process of determining a solution to your problem by making the pertinent information available to the service technicians. The IBM service technicians can start working on your solution as soon as you have completed and submitted an Electronic Service Request.

You can solve many problems without outside assistance by following the troubleshooting procedures that Lenovo provides in the online help or in the Lenovo product documentation. The Lenovo product documentation also describes the diagnostic tests that you can perform. The documentation for most systems, operating systems, and programs contains troubleshooting procedures and explanations of error messages and error codes. If you suspect a software problem, see the documentation for the operating system or program.

Appendix C. Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area.

Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service.

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.

Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties.

Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Trademarks

Lenovo, the Lenovo logo, Flex System, System x, NeXtScale System, and X-Architecture are trademarks of Lenovo in the United States, other countries, or both.

Intel and Intel Xeon are trademarks of Intel Corporation in the United States, other countries, or both.

Internet Explorer, Microsoft, and Windows are trademarks of the Microsoft group of companies.

Linux is a registered trademark of Linus Torvalds.

Other company, product, or service names may be trademarks or service marks of others.

Important Notes

Processor speed indicates the internal clock speed of the microprocessor; other factors also affect application performance.

CD or DVD drive speed is the variable read rate. Actual speeds vary and are often less than the possible maximum.

When referring to processor storage, real and virtual storage, or channel volume, KB stands for 1 024 bytes, MB stands for 1 048 576 bytes, and GB stands for 1 073 741 824 bytes.

When referring to hard disk drive capacity or communications volume, MB stands for 1 000 000 bytes, and GB stands for 1 000 000 000 bytes. Total user-accessible capacity can vary depending on operating environments.

Maximum internal hard disk drive capacities assume the replacement of any standard hard disk drives and population of all hard-disk-drive bays with the largest currently supported drives that are available from Lenovo.

Maximum memory might require replacement of the standard memory with an optional memory module.

Each solid-state memory cell has an intrinsic, finite number of write cycles that the cell can incur. Therefore, a solid-state device has a maximum number of write cycles that it can be subjected to, expressed as total bytes written (TBW). A device that has exceeded this limit might fail to respond to system-generated commands or might be incapable of being written to. Lenovo is not responsible for replacement of a device that has exceeded its maximum guaranteed number of program/erase cycles, as documented in the Official Published Specifications for the device.

Lenovo makes no representations or warranties with respect to non-Lenovo products. Support (if any) for the non-Lenovo products is provided by the third party, not Lenovo.

Some software might differ from its retail version (if available) and might not include user manuals or all program functionality.

Recycling Information

Lenovo encourages owners of information technology (IT) equipment to responsibly recycle their equipment when it is no longer needed. Lenovo offers a variety of programs and services to assist equipment owners in recycling their IT products. For information on recycling Lenovo products, go to:

http://www.lenovo.com/recycling

Particulate Contamination

Attention: Airborne particulates (including metal flakes or particles) and reactive gases acting alone or in combination with other environmental factors such as humidity or temperature might pose a risk to the device that is described in this document.

Risks that are posed by the presence of excessive particulate levels or concentrations of harmful gases include damage that might cause the device to malfunction or cease functioning altogether. This specification sets forth limits for particulates and gases that are intended to avoid such damage. The limits must not be viewed or used as definitive limits, because numerous other factors, such as temperature or moisture content of the air, can influence the impact of particulates or environmental corrosives and gaseous contaminant transfer. In the absence of specific limits that are set forth in this document, you must implement practices that maintain particulate and gas levels that are consistent with the protection of human health and safety. If Lenovo determines that the levels of particulates or gases in your environment have caused damage to the device, Lenovo may condition provision of repair or replacement of devices or parts on implementation of appropriate remedial measures to mitigate such environmental contamination. Implementation of such remedial measures is a customer responsibility..

Contaminant	Limits
Particulate	• The room air must be continuously filtered with 40% atmospheric dust spot efficiency (MERV 9) according to ASHRAE Standard 52.2[1]. • Air that enters a data center must be filtered to 99.97% efficiency or greater, using high-efficiency particulate air (HEPA) filters that meet MIL-STD-282. • The deliquescent relative humidity of the particulate contamination must be more than 60%[2]. • The room must be free of conductive contamination such as zinc whiskers.
Gaseous	• Copper: Class G1 as per ANSI/ISA 71.04-1985[3] • Silver: Corrosion rate of less than 300 Å in 30 days

[1] ASHRAE 52.2-2008 - *Method of Testing General Ventilation Air-Cleaning Devices for Removal Efficiency by Particle Size*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

[2] The deliquescent relative humidity of particulate contamination is the relative humidity at which the dust absorbs enough water to become wet and promote ionic conduction.

[3] ANSI/ISA-71.04-1985. *Environmental conditions for process measurement and control systems: Airborne contaminants*. Instrument Society of America, Research Triangle Park, North Carolina, U.S.A.

Telecommunication Regulatory Statement

This product may not be certified in your country for connection by any means whatsoever to interfaces of public telecommunications networks. Further certification may be required by law prior to making any such connection. Contact a Lenovo representative or reseller for any questions.

Electronic Emission Notices

When you attach a monitor to the equipment, you must use the designated monitor cable and any interference suppression devices that are supplied with the monitor.

Federal Communications Commission (FCC) Statement

Note: This equipment has been tested and found to comply with the limits for a Class A digital device, pursuant to Part 15 of the FCC Rules. These limits are designed to provide reasonable protection against harmful interference when the equipment is operated in a commercial environment. This equipment generates, uses, and can radiate radio frequency energy and, if not installed and used in accordance with the instruction manual, may cause harmful interference to radio communications. Operation of this equipment in a residential area is likely to cause harmful interference, in which case the user will be required to correct the interference at his own expense.

Properly shielded and grounded cables and connectors must be used to meet FCC emission limits. Lenovo is not responsible for any radio or television interference caused by using other than recommended cables and connectors or by unauthorized changes or modifications to this equipment. Unauthorized changes or modifications could void the user's authority to operate the equipment.

This device complies with Part 15 of the FCC Rules. Operation is subject to the following two conditions: (1) this device may not cause harmful interference, and (2) this device must accept any interference received, including interference that might cause undesired operation.

Industry Canada Class A Emission Compliance Statement

This Class A digital apparatus complies with Canadian ICES-003.

Avis de Conformité à la Réglementation d'Industrie Canada

Cet appareil numérique de la classe A est conforme à la norme NMB-003 du Canada.

Australia and New Zealand Class A Statement

Attention: This is a Class A product. In a domestic environment this product may cause radio interference in which case the user may be required to take adequate measures.

European Union - Compliance to the Electromagnetic Compatibility Directive

This product is in conformity with the protection requirements of EU Council Directive 2004/108/EC (until April 19, 2016) and EU Council Directive 2014/30/EU (from April 20, 2016) on the approximation of the laws of the Member States relating to electromagnetic compatibility. Lenovo cannot accept responsibility for any failure to satisfy the protection requirements resulting from a non-recommended modification of the product, including the installation of option cards from other manufacturers.

This product has been tested and found to comply with the limits for Class A equipment according to European Standards harmonized in the Directives in compliance. The limits for Class A equipment were derived for commercial and industrial environments to provide reasonable protection against interference with licensed communication equipment.

CE Lenovo, Einsteinova 21, 851 01 Bratislava, Slovakia

Warning: This is a Class A product. In a domestic environment this product may cause radio interference in which case the user may be required to take adequate measures.

Germany Class A Statement

Deutschsprachiger EU Hinweis:

Hinweis für Geräte der Klasse A EU-Richtlinie zur Elektromagnetischen Verträglichkeit

Dieses Produkt entspricht den Schutzanforderungen der EU-Richtlinie 2014/30/EU (früher 2004/108/EC) zur Angleichung der Rechtsvorschriften über die elektromagnetische Verträglichkeit in den EU-Mitgliedsstaaten und hält die Grenzwerte der Klasse A der Norm gemäß Richtlinie.

Um dieses sicherzustellen, sind die Geräte wie in den Handbüchern beschrieben zu installieren und zu betreiben. Des Weiteren dürfen auch nur von der Lenovo empfohlene Kabel angeschlossen werden. Lenovo übernimmt keine Verantwortung für die Einhaltung der Schutzanforderungen, wenn das Produkt ohne Zustimmung der Lenovo verändert bzw. wenn Erweiterungskomponenten von Fremdherstellern ohne Empfehlung der Lenovo gesteckt/eingebaut werden.

Deutschland:

Einhaltung des Gesetzes über die elektromagnetische Verträglichkeit von Betriebsmittein

Dieses Produkt entspricht dem „Gesetz über die elektromagnetische Verträglichkeit von Betriebsmitteln“ EMVG (früher „Gesetz über die elektromagnetische Verträglichkeit von Geräten“). Dies ist die Umsetzung der EU-Richtlinie 2014/30/EU (früher 2004/108/EC) in der Bundesrepublik Deutschland.

Zulassungsbescheinigung laut dem Deutschen Gesetz über die elektromagnetische Verträglichkeit von Betriebsmitteln, EMVG vom 20. Juli 2007 (früher Gesetz über die elektromagnetische Verträglichkeit von Geräten), bzw. der EMV EU Richtlinie 2014/30/EU (früher 2004/108/EC), für Geräte der Klasse A.

Dieses Gerät ist berechtigt, in Übereinstimmung mit dem Deutschen EMVG das EG-Konformitätszeichen - CE - zu führen. Verantwortlich für die Konformitätserklärung nach Paragraf 5 des EMVG ist die Lenovo (Deutschland) GmbH, Meitnerstr. 9, D-70563 Stuttgart.

Informationen in Hinsicht EMVG Paragraf 4 Abs. (1) 4:

Das Gerät erfüllt die Schutzanforderungen nach EN 55024 und EN 55022 Klasse A.

Nach der EN 55022: „Dies ist eine Einrichtung der Klasse A. Diese Einrichtung kann im Wohnbereich Funkstörungen verursachen; in diesem Fall kann vom Betreiber verlangt werden, angemessene Maßnahmen durchzuführen und dafür aufzukommen."

Nach dem EMVG: „Geräte dürfen an Orten, für die sie nicht ausreichend entstört sind, nur mit besonderer Genehmigung des Bundesministers für Post und Telekommunikation oder des Bundesamtes für Post und Telekommunikation betrieben werden. Die Genehmigung wird erteilt, wenn keine elektromagnetischen Störungen zu erwarten sind." (Auszug aus dem EMVG, Paragraph 3, Abs. 4). Dieses Genehmigungsverfahrenist nach Paragraph 9 EMVG in Verbindung mit der entsprechenden Kostenverordnung (Amtsblatt 14/93) kostenpflichtig.

Anmerkung: Um die Einhaltung des EMVG sicherzustellen sind die Geräte, wie in den Handbüchern angegeben, zu installieren und zu betreiben.

Japan VCCI Class A Statement

この装置は、クラス A 情報技術装置です。この装置を家庭環境で使用すると電波妨害を引き起こすことがあります。この場合には使用者が適切な対策を講ずるよう要求されることがあります。 VCCI–A

This is a Class A product based on the standard of the Voluntary Control Council for Interference (VCCI). If this equipment is used in a domestic environment, radio interference may occur, in which case the user may be required to take corrective actions.

Japan Electronics and Information Technology Industries Association (JEITA) Statement

高調波ガイドライン適合品

Japan Electronics and Information Technology Industries Association (JEITA) Confirmed Harmonics Guidelines (products less than or equal to 20 A per phase)

高調波ガイドライン準用品

Japan Electronics and Information Technology Industries Association (JEITA) Confirmed Harmonics Guidelines with Modifications (products greater than 20 A per phase).

Korea Communications Commission (KCC) Statement

이 기기는 업무용(A급)으로 전자파적합기기로
서 판매자 또는 사용자는 이 점을 주의하시기
바라며, 가정외의 지역에서 사용하는 것을 목
적으로 합니다.

This is electromagnetic wave compatibility equipment for business (Type A). Sellers and users need to pay attention to it. This is for any areas other than home.

Russia Electromagnetic Interference (EMI) Class A statement

ВНИМАНИЕ! Настоящее изделие относится к классу А.
В жилых помещениях оно может создавать радиопомехи, для
снижения которых необходимы дополнительные меры

People's Republic of China Class A electronic emission statement

中华人民共和国“A类”警告声明

声 明

此为A级产品，在生活环境中，该产品可能会造成无线电干扰。在这种情况下，
可能需要用户对其干扰采取切实可行的措施。

Taiwan Class A compliance statement

警告使用者:
這是甲類的資訊產品，在
居住的環境中使用時，可
能會造成射頻干擾，在這
種情況下，使用者會被要
求採取某些適當的對策。

Index

Symbols

Numerics

A

B

C

D

E

F

G

H

I

J

K

L

M

N

S

T

U

V

W

Lenovo

Part Number: 00MY375

Printed in USA

(IP) P/N: 00MY375