BLADEOS™ 6.5

Application Guide

RackSwitch™ G8000

Part Number: BMD00215, September 2010

BLADE
NETWORK TECHNOLOGIES
2051 Mission College Blvd.
Santa Clara, CA 95054
www.bladenetwork.net

Originated in the USA.

Contents

Preface

The *BLADEOS 6.5 Application Guide* describes how to configure and use the BLADEOS 6.5 software on the RackSwitch G8000 (referred to as G8000 throughout this document). For documentation on installing the switch physically, see the *Installation Guide* for your G8000.

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 BLADEOS 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 BLADEOS products with the basics of switch management. This part includes the following chapters:

- Chapter 1, "Switch Administration," describes how to access the G8000 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.

Part 2: Securing the Switch

- Chapter 3, "Securing Administration," describes methods for using Secure Shell for administration connections, and configuring end-user access control.
- Chapter 4, "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 5, "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 G8000 that connect to blade servers.
- Chapter 6, "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 7, "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 8, "Ports and Trunking," describes how to group multiple physical ports together to aggregate the bandwidth between large-scale network devices.
- Chapter 9, "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 10, "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 11, "Stacking," describes how to combine multiple switches into a single, aggregate switch entity.
- Chapter 12, "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 13, "VMready," discusses virtual machine (VM) support on the G8000.

Part 5: IP Routing

- Chapter 14, “Basic IP Routing,” describes how to configure the G8000 for IP routing using IP subnets, BOOTP, and DHCP Relay.
- Chapter 15, “Internet Protocol Version 6,” describes how to configure the G8000 for IPv6 host management.
- Chapter 16, “Routing Information Protocol,” describes how the BLADEOS software implements standard Routing Information Protocol (RIP) for exchanging TCP/IP route information with other routers.
- Chapter 17, “Internet Group Management Protocol,” describes how the BLADEOS software implements IGMP Snooping or IGMP Relay to conserve bandwidth in a multicast-switching environment.
- Chapter 18, “Border Gateway Protocol,” describes Border Gateway Protocol (BGP) concepts and features supported in BLADEOS.
- Chapter 19, “OSPF,” describes key Open Shortest Path First (OSPF) concepts and their implemented in BLADEOS, and provides examples of how to configure your switch for OSPF support.
- Chapter 20, “Protocol Independent Multicast,” describes how multicast routing can be efficiently accomplished using the Protocol Independent Multicast (PIM) feature.

Part 6: High Availability Fundamentals

- Chapter 21, “Basic Redundancy,” describes how the G8000 supports redundancy through stacking, trunking, Active Multipass Protocol (AMP), and hotlinks.
- Chapter 22, “Layer 2 Failover,” describes how the G8000 supports high-availability network topologies using Layer 2 Failover.
- Chapter 23, “Virtual Router Redundancy Protocol,” describes how the G8000 supports high-availability network topologies using Virtual Router Redundancy Protocol (VRRP).

Part 7: Network Management

- Chapter 24, “Link Layer Discovery Protocol,” describes how Link Layer Discovery Protocol helps neighboring network devices learn about each others’ ports and capabilities.
- Chapter 25, “Simple Network Management Protocol,” describes how to configure the switch for management through an SNMP client.

Part 8: Monitoring

- Chapter 26, "Remote Monitoring," describes how to configure the RMON agent on the switch, so that the switch can exchange network monitoring data.
- Chapter 27, "sFLOW, described how to use the embedded sFlow agent for sampling network traffic and providing continuous monitoring information to a central sFlow analyzer.
- Chapter 28, "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.

Additional References

Additional information about installing and configuring the G8000 is available in the following guides:

- *RackSwitch G8000 Installation Guide*
- *BLADEOS 6.5 Command Reference*
- *BLADEOS 6.5 ISCLI Reference Guide*
- *BLADEOS 6.5 BBI 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.	View the `readme.txt` file.
	It also depicts on-screen computer output and prompts.	`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.	To establish a Telnet session, enter: `host#` **`telnet`** *<IP address>*
	This also shows book titles, special terms, or words to be emphasized.	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.

How to Get Help

If you need help, service, or technical assistance, call BLADE Network Technologies Technical Support:

US toll free calls: 1-800-414-5268

International calls: 1-408-834-7871

You also can visit our web site at the following address:

`http://www.bladenetwork.net`

Click the **Support** tab.

The warranty card received with your product provides details for contacting a customer support representative. If you are unable to locate this information, please contact your reseller. Before you call, prepare the following information:

- Serial number of the switch unit
- Software release version number
- Brief description of the problem and the steps you have already taken
- Technical support dump information (# **`show tech-support`**)

Part 1: Getting Started

CHAPTER 1

Switch Administration

Your RackSwitch G8000 (G8000) 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 BLADEOS switching software included in the G8000 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

BLADEOS 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:

- 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 or HP OpenView

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 *RackSwitch G8000 Installation Guide*).

Command Line Interface

The BLADEOS Command Line Interface (CLI) provides a simple, direct method for switch administration. Using a basic terminal, you are presented with an organized hierarchy of menus, each with logically-related sub-menus and commands. These allow you to view detailed information and statistics about the switch, and to perform any necessary configuration and switch software maintenance. For example:

```
[Main Menu]
      info    - Information Menu
      stats   - Statistics Menu
      cfg     - Configuration Menu
      oper    - Operations Command Menu
      boot    - Boot Options Menu
      maint   - Maintenance Menu
      diff    - Show pending config changes  [global command]
      apply   - Apply pending config changes [global command]
      save    - Save updated config to FLASH [global command]
      revert  - Revert pending or applied changes [global command]
      exit    - Exit  [global command, always available]
>> #
```

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

- Serial connection via the serial port on the G8000 (this option is always available)
- Telnet connection over the network
- SSH connection over the network

Browser-Based Interface

The Browser-based Interface (BBI) provides access to the common configuration, management and operation features of the G8000 through your Web browser.

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

Establishing a Connection

The factory default settings permit initial switch administration through *only* the built-in serial port. All other forms of access require additional switch configuration before they can be used.

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 34), 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.

To manage the switch using Telnet, SNMP, or a Web browser, you must configure an IP interface.

When a DHCP server is present in the local network for the switch, the DHCP server will be used to configures the IP interface. However, if the switch fails to renew the address obtained through DHCP, the following factory configured settings will be used for IP interface 1:

IPv4 address:	192.168.1.211
Mask:	255.255.255.0
Gateway:	192.168.1.255
DHCP:	`enabled`

If you manually configure a static IP address, DHCP is disabled. If you manually enable DHCP, the interface will be configured by the DHCP server.

To access the switch, the following IP parameters must be configured:

1. Log on to the switch.
2. Enter IP interface mode.

```
RS G8000> enable
RS G8000# configure terminal
RS G8000(config)# interface ip <IP interface number>
```

3. Configure the management IP interface/mask.
   - Using IPv4:

   ```
   RS G8000(config-ip-if)# ip address <management interface IPv4 address>
   RS G8000(config-ip-if)# ip netmask <IPv4 subnet mask>
   ```

   - Using IPv6:

   ```
   RS G8000(config-ip-if)# ipv6 address <management interface IPv6 address>
   RS G8000(config-ip-if)# ipv6 prefixlen <IPv6 prefix length>
   ```

4. Configure the VLAN, and enable the interface.

   ```
   RS G8000(config-ip-if)# vlan 1
   RS G8000(config-ip-if)# enable
   RS G8000(config-ip-if)# exit
   ```

5. Configure the default gateway.
   - If using IPv4:

   ```
   RS G8000(config)# ip gateway <gateway number> address <IPv4 address>
   RS G8000(config)# ip gateway <gateway number> enable
   ```

   - If using IPv6:

   ```
   RS G8000(config)# ip gateway6 <gateway number> address <IPv6 address>
   RS G8000(config)# ip gateway6 <gateway number> enable
   ```

Once you configure the IP address and you have an existing network connection, you can use the Telnet program from an external management station to access and control the switch. Once the default gateway is enabled, the management station and your switch do not need to be on the same IP subnet.

The G8000 supports a menu-based command-line interface (CLI) as well as an industry standard command-line interface (ISCLI) that you can use to configure and control the switch over the network using the Telnet program. You can use the CLI or ISCLI to perform many basic network management functions. In addition, you can configure the switch for management using an SNMP-based network management system or a Web browser.

For more information, see the documents listed in "Additional References" on page 20.

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 enabled. Use the following commands (available on the console only) to disable or re-enable Telnet access:

```
RS G8000(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 36.

Using Secure Shell

Although a remote network administrator can manage the configuration of a G8000 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: RSA
- Encryption: 3DES-CBC, DES
- 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 BLADEOS implementation of SSH supports both versions 1.5 and 2.0 and supports SSH client version 1.5 - 2.x.

Using SSH to Access the Switch

By default, the SSH feature is disabled. For information on enabling and using SSH for switch access, see "Secure Shell and Secure Copy" on page 65.

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>
```

If SecurID authentication is required, use the following command:

```
# ssh -1 ace <switch IP address>
```

You will then be prompted to enter a password as explained "Switch Login Levels" on page 36.

Using a Web Browser

The switch provides a Browser-Based Interface (BBI) for accessing the common configuration, management and operation features of the G8000 through your Web browser.

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

You can also 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>*).

Configuring HTTP Access to the BBI

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

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

```
RS G8000(config)# access http enable            (Enable HTTP access)
   -or-
RS G8000(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:

```
RS G8000(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 disabled on the switch. To enable BBI Access via HTTPS, use the following command:

   ```
   RS G8000(config)# access https enable
   ```

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

   To change the HTTPS Web server port number from the default port 443, use the following command:

   ```
   RS G8000(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.

   ```
   RS G8000(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. In order to save the certificate so that it is retained beyond reboot or power cycles, use the following command:

   ```
   RS G8000(config)# access https save-certificate
   ```

When a client (e.g. 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 *BLADEOS 6.5 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 information on using the BBI, refer to the *BLADEOS 6.5 BBI Quick Guide*.

Using Simple Network Management Protocol

BLADEOS 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 or HP-OpenView.

Note – SNMP read and write functions are enabled by default. For best security practices, if SNMP is not needed for your network, it is recommended that you disable these functions prior to connecting the switch to the network.

To access the SNMP agent on the G8000, the read and write community strings on the SNMP manager should be configured to match those on the switch. The default read community string on the switch is `public` and the default write community string is `private`.

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

```
RS G8000(config)# snmp-server read-community <1-32 characters>
    -and-
RS G8000(config)# snmp-server write-community <1-32 characters>
```

The SNMP manager should be able to reach 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 commands:

```
RS G8000(config)# snmp-server trap-src-if <trap source IP interface>
RS G8000(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 347.

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 G8000 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 four BOOTP/DHCP servers. In addition to the four global BOOTP/DHCP servers, up to four 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 G8000 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 G8000.

Global BOOTP Relay Agent Configuration

To enable the G8000 to be a BOOTP (or DHCP) forwarder, enable the BOOTP relay feature, configure up to four global BOOTP server IPv4 addresses on the switch, and enable BOOTP relay on the interface(s) on which the client requests are expected.

Generally, you should configure BOOTP for the switch IP interface that is closest to the client, so that the BOOTP server knows from which IPv4 subnet the newly allocated IPv4 address should come.

In the G8000 implementation, there are no primary or secondary BOOTP servers. The client request is forwarded to all the global BOOTP servers configured on the switch (if no domain-specific servers are configured). The use of multiple servers provide failover redundancy. However, no health checking is supported.

1. Use the following commands to configure global BOOTP relay servers:

```
RS G8000(config)# ip bootp-relay enable
RS G8000(config)# ip bootp-relay server <1-4> address <IPv4 address>
```

2. Enable BOOTP relay on the appropriate IP interfaces.

BOOTP/DHCP Relay functionality may be assigned on a per-interface basis using the following commands:

```
RS G8000(config)# interface ip <interface number>
RS G8000(config-ip-if)# relay
RS G8000(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:

```
RS G8000(config)# ip bootp-relay bcast-domain <1-10> vlan <VLAN number>
RS G8000(config)# ip bootp-relay bcast-domain <1-10> server <1-4>
    address <IPv4 address>
RS G8000(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 (see Step 2 in page 35).

Switch Login Levels

To enable better switch management and user accountability, three levels or *classes* of user access have been implemented on the G8000. 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 G8000. 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 G8000. 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 G8000. 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 surnames and passwords. Once you are connected to the switch via local Telnet, 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 61.

Table 2 User Access Levels

User Account	Password	Description and Tasks Performed
`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.
`oper`	`oper`	The Operator manages all functions of the switch. The Operator can reset ports, except the management ports.
`admin`	`admin`	The superuser Administrator has complete access to all menus, information, and configuration commands on the G8000, including the ability to change both the user and administrator passwords.

Note – With the exception of the "admin" user, access to each user level can be disabled by setting the password to an empty value.

Setup vs. the Command Line

Once the administrator password is verified, you are given complete access to the switch. If the switch is still set to its factory default configuration, the system will ask whether you wish to run Setup (see "Initial Setup" on page 39"), a utility designed to help you through the first-time configuration process. If the switch has already been configured, the command line is displayed instead.

CHAPTER 2

Initial Setup

To help with the initial process of configuring your switch, the BLADEOS 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.

Whenever you log in as the system administrator under the factory default configuration, you are asked whether you wish to run the Setup utility. Setup can also be activated manually from the command line interface any time after login.

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 below.

```
Enter Password:
```

2. Enter **`admin`** as the default administrator password.

   If the factory default configuration is detected, the system prompts:

```
RackSwitch G8000
18:44:05 Wed Jan 3, 2009

The switch is booted with factory default configuration.
To ease the configuration of the switch, a "Set Up" facility which
will prompt you with those configuration items that are essential to
the operation of the switch is provided.
Would you like to run "Set Up" to configure the switch? [y/n]:
```

Note – If the default `admin` login is unsuccessful, or if the administrator Main Menu appears instead, the system configuration has probably been changed from the factory default settings. If desired, return the switch to its factory default configuration.

---

3. Enter **`y`** to begin the initial configuration of the switch, or n to bypass the Setup facility.

Stopping and Restarting Setup Manually

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:

```
# /cfg/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 *BLADEOS 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 [2009]:
   ```

   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, 2009.
   ```

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, 2009.
   ```

8. 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 VLAN tagging 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 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 44.

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 VLAN tagging config (tagged port can be a member of multiple VLANs)
Current VLAN tag support:              disabled
Enter new VLAN tag 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 (INT1-14, MGT1-2, EXT1-48):
```

   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 45.

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 45.

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 or port alias, 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 Spanning Tree Group index [1-127]:
```

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 219|.

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 RackSwitch G8000 (G8000). 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).

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 47.

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: (1-4)
```

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 47.

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 G8000, 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 61.

Optional Setup for Telnet Support

Note – This step is optional. Perform this procedure only if you are planning on connecting to the G8000 through a remote Telnet connection.

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

```
>> # /cfg/sys/access/tnet
```

2. Apply and save the configuration(s).

```
>> System# apply
>> System# save
```

Part 2: Securing the Switch

CHAPTER 3

Securing Administration

Secure switch management is needed for environments that perform significant management functions across the Internet. Common functions for secured management are described in the following sections:

- "Secure Shell and Secure Copy" on page 53
- "End User Access Control" on page 60

Note – SNMP read and write functions are enabled by default. For best security practices, if SNMP is not needed for your network, it is recommended that you disable these functions prior to connecting the switch to the network (see "Using Simple Network Management Protocol" on page 33).

Secure Shell and Secure Copy

Because using Telnet does not provide a secure connection for managing a G8000, Secure Shell (SSH) and Secure Copy (SCP) features have been included for G8000 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 G8000 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 G8000, 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 BLADEOS 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 features are disabled by default. To change the SSH/SCP settings, using the following procedures.

To Enable or Disable the SSH Feature

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

```
RS G8000(config)# [no] ssh enable
```

To Enable or Disable SCP Apply and Save

Enter the following commands from the switch CLI to enable the SCP `putcfg_apply` and `putcfg_apply_save` commands:

```
RS G8000(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`):

```
RS G8000(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 client commands. The examples below use 205.178.15.157 as the IP address of a sample switch.

To Log In to the Switch

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, SecurID (via RADIUS or TACACS+ for SSH only—does not apply to SCP)

Generating RSA Host and Server Keys for SSH Access

To support the SSH server feature, two sets of RSA keys (host and server keys) are required. The host key is 1024 bits and is used to identify the G8000. The server key is 768 bits and is used to make it impossible to decipher a captured session by breaking into the G8000 at a later time.

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

To configure RSA host and server keys, first connect to the G8000 through the console port (commands are not available via external Telnet connection), and enter the following commands to generate them manually.

```
RS G8000(config)# ssh generate-host-key
RS G8000(config)# ssh generate-server-key
```

When the switch reboots, it will retrieve the host and server keys from the FLASH memory. If these two keys are not available in the flash and if the SSH server feature is enabled, the switch automatically generates them during the system reboot. This process may take several minutes to complete.

The switch can also automatically regenerate the RSA server key. To set the interval of RSA server key autogeneration, use this command:

```
RS G8000(config)# ssh interval <number of hours (0-24)>
```

A value of 0 (zero) denotes that RSA server key autogeneration is disabled. When greater than 0, the switch will autogenerate the RSA server key every specified interval; however, RSA server key generation is skipped if the switch is busy doing other key or cipher generation when the timer expires.

Note – 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.

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.

SecurID Support

SSH/SCP can also work with SecurID, a token card-based authentication method. The use of SecurID requires the interactive mode during login, which is not provided by the SSH connection.

Note – There is no SNMP or Browser-Based Interface (BBI) support for SecurID because the SecurID server, ACE, is a one-time password authentication and requires an interactive session.

Using SecurID with SSH

Using SecurID with SSH involves the following tasks.

- To log in using SSH, use a special username, “ace,” to bypass the SSH authentication.
- After an SSH connection is established, you are prompted to enter the username and password (the SecurID authentication is being performed now).
- Provide your username and the token in your SecurID card as a regular Telnet user.

Using SecurID with SCP

Using SecurID with SCP can be accomplished in two ways:

- Using a RADIUS server to store an administrator password.

  You can configure a regular administrator with a fixed password in the RADIUS server if it can be supported. A regular administrator with a fixed password in the RADIUS server can perform both SSH and SCP with no additional authentication required.

- Using an SCP-only administrator password.

  Set the SCP-only administrator password (**`ssh scp-password`**) to bypass checking SecurID.

  An SCP-only administrator’s password is typically used when SecurID is not used. For example, it can be used in an automation program (in which the tokens of SecurID are not available) to back up (download) the switch configurations each day.

Note – The SCP-only administrator’s password must be different from the regular administrator’s password. If the two passwords are the same, the administrator using that password will not be allowed to log in as an SSH user because the switch will recognize him as the SCP-only administrator. The switch will only allow the administrator access to SCP commands.

End User Access Control

BLADEOS 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 10 user IDs are supported on the switch.
- BLADEOS 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 G8000. Also note that the password change command only modifies only the user 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 for end users can be up to 128 characters in length for TACACS, RADIUS, Telnet, SSH, Console, and Web access.

Strong Passwords

The administrator can require use of Strong Passwords for users to access the G8000. Strong Passwords enhance security because they make password guessing more difficult.

The following rules apply when Strong Passwords are enabled:

- Each passwords must be 8 to 14 characters
- Within the first 8 characters, the password:
    - must have at least one number or one symbol
    - must have both upper and lower case letters
    - cannot be the same as any four previously used passwords

The following are examples of strong passwords:

- 1234AbcXyz
- Super+User
- Exo1cet2

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.

Use the Strong Password commands to configure Strong Passwords.

```
>> # access user strong-password enable
```

User Access Control

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

Setting up User IDs

Up to 10 user IDs can be configured. Use the following commands to define any user name and set the user password at the resulting prompts:

```
RS G8000(config)# access user 1 name <1-8 characters>
RS G8000(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 3 on page 66.

To change the user’s level, select one of the following options:

```
RS G8000(config)# access user 1 level {user|operator|administrator}
```

Validating a User’s Configuration

```
RS G8000# 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.

```
RS G8000(config)# [no] access user 1 enable
```

Listing Current Users

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

```
RS G8000# show access user

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

Current User ID table:
1: name jane   , ena, cos user    , password valid, online 1 session
2: name john   , ena, cos user    , password valid, online 2 sessions
```

Logging into 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 COS established for the end user account.

CHAPTER 4
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:

- "RADIUS Authentication and Authorization" on page 63
- "TACACS+ Authentication" on page 67
- "LDAP Authentication and Authorization" on page 70

Note – BLADEOS 6.5 does not support IPv6 for RADIUS, TACACS+ or LDAP.

RADIUS Authentication and Authorization

BLADEOS 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 G8000—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 switch.

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

```
RS G8000(config)# radius-server primary-host 10.10.1.1
RS G8000(config)# radius-server secondary-host 10.10.1.2
RS G8000(config)# radius-server enable
```

2. Configure the RADIUS secret.

```
RS G8000(config)# radius-server primary-host 10.10.1.1 key
        <1-32 character secret>
RS G8000(config)# radius-server secondary-host 10.10.1.2 key
        <1-32 character secret>
```

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

   The well-known port for RADIUS is 1812.

```
RS G8000(config)# radius-server port <UDP port number>
```

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

```
RS G8000(config)# radius-server retransmit 3
RS G8000(config)# radius-server timeout 5
```

RADIUS Authentication Features in BLADEOS

BLADEOS supports the following RADIUS authentication features:

- Supports RADIUS client on the switch, based on the protocol definitions in RFC 2138 and RFC 2866.
- Allows RADIUS secret password up to 32 bytes and less than 16 octets.
- 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:

  ```
  RS G8000# 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 in 1-3 retries. The switch will also automatically retry connecting to the RADIUS server before it declares the server down.

- Supports user-configurable RADIUS application port.
  The default is 1812/UDP-based on RFC 2138. Port 1645 is also supported.
- 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 3 can be defined in the RADIUS server dictionary file.

Table 3 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	The Operator manages all functions of the switch. The Operator can reset ports.	`oper`
Administrator	The super-user Administrator has complete access to all commands, information, and configuration commands on the switch, including the ability to change both the user and administrator passwords.	`admin`

RADIUS Attributes for BLADEOS 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 an option to allow *secure backdoor* access via Telnet/SSH/BBI. Secure backdoor provides switch access when the RADIUS servers cannot be reached. You always can access the switch via the console port, by using `noradius` and the administrator password, whether secure backdoor is enabled or not.

Note – To obtain the RADIUS backdoor password for your G8000, contact Technical Support.

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 G8000 user privileges levels:

Table 4 BLADEOS-proprietary Attributes for RADIUS

User Name/Access	User-Service-Type	Value
User	*Vendor-supplied*	255
Operator	*Vendor-supplied*	252
Admin	*Vendor-supplied*	6

TACACS+ Authentication

BLADEOS supports authentication and authorization with networks using the Cisco Systems TACACS+ protocol. The G8000 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 G8000 through a data 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 63.

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.

TACACS+ Authentication Features in BLADEOS

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. BLADEOS 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 BLADEOS management access levels is shown in Table 5. The authorization levels must be defined on the TACACS+ server.

Table 5 Default TACACS+ Authorization Levels

BLADEOS User Access Level	TACACS+ level
`user`	0
`oper`	3
`admin`	6

Alternate mapping between TACACS+ authorization levels and BLADEOS management access levels is shown in Table 6. Use the following command to set the alternate TACACS+ authorization levels.

```
RS G8000(config)# tacacs-server privilege-mapping
```

Table 6 Alternate TACACS+ Authorization Levels

BLADEOS User Access Level	TACACS+ level
`user`	0 - 1
`oper`	6 - 8
`admin`	14 - 15

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. The administrator has an option to allow *secure backdoor* access via Telnet/SSH. Secure backdoor provides switch access when the TACACS+ servers cannot be reached. You always can access the switch via the console port, by using `notacacs` and the administrator password, whether secure backdoor is enabled or not.

Note – To obtain the TACACS+ backdoor password for your G8000, contact Technical Support.

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 G8000 supports the following TACACS+ accounting attributes:

- protocol (console/Telnet/SSH/HTTP/HTTPS)
- 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 **Logout** button on the browser is clicked.

Command Authorization and Logging

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

```
RS G8000(config)# tacacs-server command-authorization
```

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

```
RS G8000(config)# tacacs-server command-logging
```

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

```
authorization request, cmd=shell, cmd-arg=interface ip
accounting request, cmd=shell, cmd-arg=interface ip
authorization request, cmd=shell, cmd-arg=enable
accounting request, cmd=shell, cmd-arg=enable
```

Configuring TACACS+ Authentication on the Switch

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

```
RS G8000(config)# tacacs-server primary-host 10.10.1.1
RS G8000(config)# tacacs-server secondary-host 10.10.1.2
RS G8000(config)# tacacs-server enable
```

2. Configure the TACACS+ secret and second secret.

```
RS G8000(config)# tacacs-server primary-host 10.10.1.1 key
    <1-32 character secret>
RS G8000(config)# tacacs-server secondary-host 10.10.1.2 key
    <1-32 character secret>
```

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

   The well-known port for TACACS+ is 49.

```
RS G8000(config)# tacacs-server port <TCP port number>
```

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

```
RS G8000(config)# tacacs-server retransmit 3
RS G8000(config)# tacacs-server timeout 5
```

LDAP Authentication and Authorization

BLADEOS 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

G8000 user groups and user accounts must reside within the same domain. On the LDAP server, configure the domain to include G8000 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 G8000, as follows:

  - `admin`
  - `oper`
  - `user`

Configuring LDAP Authentication on the Switch

1. Turn LDAP authentication on, then configure the IPv4 addresses of the Primary and Secondary LDAP servers.

```
>> # ldap-server enable
>> # ldap-server primary-host 10.10.1.1
>> # ldap-server secondary-host 10.10.1.2
```

2. Configure the domain name.

```
>> # ldap-server domain <ou=people,dc=my-domain,dc=com>
```

3. You may change the default TCP port number used to listen to LDAP (optional).

   The well-known port for LDAP is 389.

```
>> # ldap-server port <1-65000>
```

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

```
>> # ldap-server retransmit 3
>> # ldap-server timeout 10
```

CHAPTER 5

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 RackSwitch G8000 (G8000) that connect to blade servers.

The following topics are discussed in this section:

Extensible Authentication Protocol over LAN

BLADEOS 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:

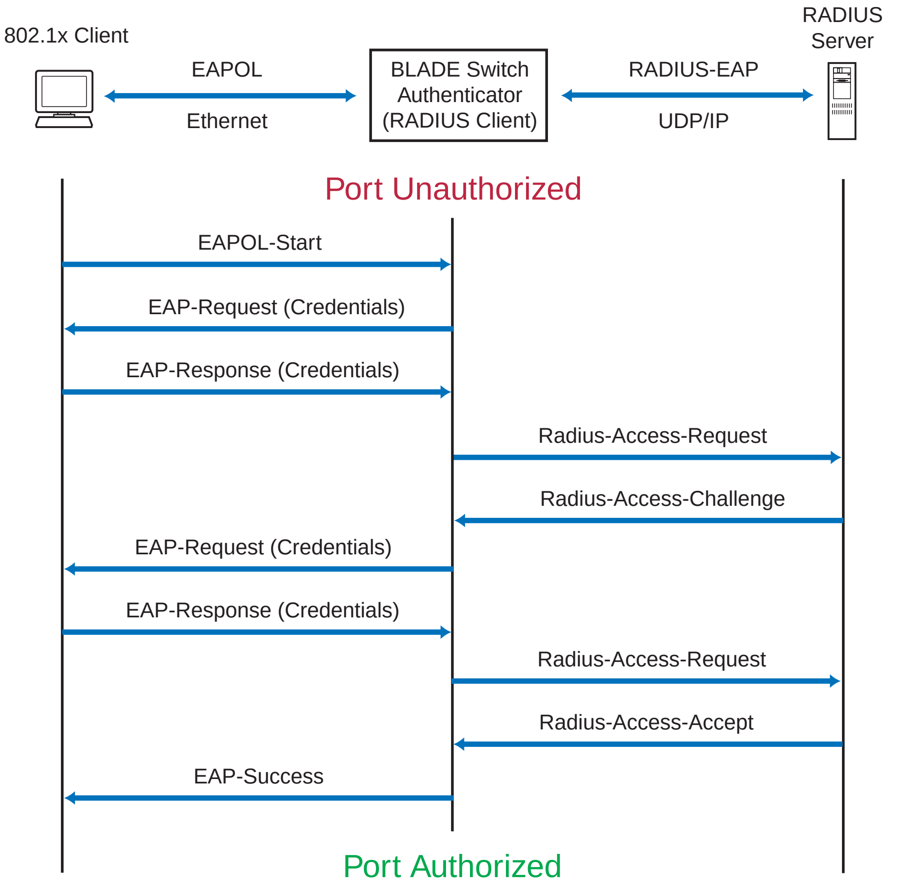
- Supplicant or Client
  The Supplicant is a device that requests network access and provides the required credentials (user name and password) to the Authenticator and the Authenticator Server.
- Authenticator
  The Authenticator enforces authentication and controls access to the network. The Authenticator 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 G8000 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 G8000 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 G8000 authenticator, while RADIUS-EAP messages are exchanged between the G8000 authenticator and the RADIUS server.

Authentication is initiated by one of the following methods:

- The G8000 authenticator sends an EAP-Request/Identity packet to the client
- The client sends an EAPOL-Start frame to the G8000 authenticator, which responds with an EAP-Request/Identity frame.

The client confirms its identity by sending an EAP-Response/Identity frame to the G8000 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 G8000 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 G8000 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 G8000 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.

Use the 802.1X global configuration commands (`dot1x`) to configure 802.1X authentication for all ports in the switch. Use the 802.1X port commands to configure a single port.

Guest VLAN

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

```
RS G8000(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 7 lists 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 7 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 7 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 or stacking) 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 G8000 is connected to another G8000, 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 6

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.

BLADEOS 6.5 supports the following ACLs:

- Regular ACLs

  Up to 512 ACLs are supported for networks that use IPv4 addressing. Regular ACLs are configured using the following ISCLI command path:

  ```
  RS G8000(config)# access-control list <Regular ACL number> ?
  ```

- IPv6 ACLs
- Up to 128 ACLs are supported for networks that use IPv6 addressing. IPv6 ACLs are configured using the following ISCLI command path:VLAN Maps (VMaps)

  ```
  RS G8000(config)# access-control list6 <IPv6 ACL number> ?
  ```

  Up to 128 VLAN Maps are supported for attaching filters to VLANs rather than ports. See “VLAN Maps” on page 92 for details.

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, IPv6 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, IP, IPv6, 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 8:

Table 8 Well-Known Protocol Types

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

- IPv6 header options (for IPv6 ACLs only)
  - Source IPv6 address and prefix length
  - Destination IPv6 address and prefix length
  - Next Header value
  - Flow Label value
  - Traffic Class value

- TCP/UDP header options (for all ACLs)
  - TCP/UDP application source port as shown in Table 9

Table 9 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 9
  - TCP/UDP flag value as shown in Table 10

Table 10 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, IPv6)
- 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. G8000 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 ACLs to a port, use the following IP Interface Mode commands:

```
RS G8000(config)# interface port <port>
RS G8000(config-ip)# access-control list <Regular ACL number>
RS G8000(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.

Note – When IPv6 ACLs are applied to a port, some regular ACLs are restricted from being applied to the same port. Only regular ACLs 1 through 256 may be applied to ports that also use IPv6 ACLs.

To create and assign ACLs in groups, see “ACL Groups” on page 86.

ACL Order of Precedence

When multiple ACLs are assigned to a port, the order in which the ACLs are applied to port traffic (or whether they are applied at all) depends on the following factors:

- The precedence group in which the ACL resides;
- The ACL number;
- Whether a prior ACL in the precedence group is also matched;
- And whether the ACL action is compatible with preceding ACLs.

ACLs are automatically divided into precedence groups as follows:

Precedence Group 1 includes ACL 1–128.
Precedence Group 2 includes ACL 129–256.
Precedence Group 3 includes ACL 257–384.
Precedence Group 4 includes ACL 385–512.

The switch processes each precedence group in numeric sequence; Precedence group 1 is evaluated first, followed by precedence group 2, and so on.

Within each precedence group, ACLs assigned to the port are processed in numeric sequence, based on 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 within precedence group 1.

For each precedence group, only the first assigned ACL that matches the port traffic is considered. If multiple ACLs in the precedence group match the traffic, only the one with the lowest ACL number is considered. The others in the precedence group are ignored.

One ACL match from each precedence group is permitted, meaning that up to four ACL matches may be considered for action: one from precedence group 1, one from precedence group 2, and so on.

Of the matching ACLs permitted, each configured ACL action is applied in sequence, based on ACL number, with the lowest-numbered ACL's action applied first. If an ACL action contradicts a preceding ACL (one with a lower ACL number), the action of the higher-numbered ACL is ignored.

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 G8000 supports up to 512 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 G8000 supports up to 512 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 85). 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 command:

```
RS G8000(config-if)# access-control group <ACL group number>
RS G8000(config-if)# exit
```

ACL Metering and Re-Marking

You can define a profile for the aggregate traffic flowing through the G8000 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 85).

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 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:

```
RS G8000(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 84, or "ACL Groups" on page 86 and "Assigning ACL Groups to a Port" on page 87).

- For VMaps (see "VLAN Maps" on page 92):

```
RS G8000(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:

```
RS G8000(config)# access-control list <ACL number> statistics
```

ACL Configuration Examples

ACL Example 1

Use this configuration to block traffic to a specific host. All traffic that ingresses on port 1 is denied if it is destined for the host at IP address 100.10.1.1

1. Configure an Access Control List.

```
RS G8000(config)# access-control list 1 ipv4 destination-ip-address
    100.10.1.1
RS G8000(config)# access-control list 1 action deny
```

2. Add ACL 1 to port EXT1.

```
RS G8000(config)# interface port 1
RS G8000(config-if)# access-control list 1
RS G8000(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 2 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.

```
RS G8000(config)# access-control list 2 ipv4 source-ip-address
    100.10.1.0 255.255.255.0
RS G8000(config)# access-control list 2 ipv4 destination-ip-address
    200.20.2.2 255.255.255.255
RS G8000(config)# access-control list 1 action deny
```

2. Add ACL 2 to port EXT2.

```
RS G8000(config)# interface port 2
RS G8000(config-if)# access-control list 2
RS G8000(config-if)# exit
```

ACL Example 3

Use this configuration to block traffic from a specific IPv6 source address. All traffic that ingresses in port 2 with source IP from class 2001:0:0:5:0:0:0:2/128 is denied.

1. Configure an Access Control List.

```
RS G8000(config)# access-control list6 3 ipv6 source-address
    2001:0:0:5:0:0:0:2 128
RS G8000(config)# access-control list6 3 action deny
```

2. Add ACL 2 to port EXT2.

```
RS G8000(config)# interface port 2
RS G8000(config-if)# access-control list6 3
RS G8000(config-if)# exit
```

ACL Example 4

This configuration blocks traffic from a network that is destined for a specific egress port. All traffic that ingresses port 1 from the network 100.10.1.0/24 and is destined for port 3 is denied.

1. Configure an Access Control List.

```
RS G8000(config)# access-control list 4 ipv4 source-ip-address
    100.10.1.0 255.255.255.0
RS G8000(config)# access-control list 4 egress-port 3
RS G8000(config)# access-control list 4 action deny
```

2. Add ACL 4 to port EXT1.

```
RS G8000(config)# access-control list 4 ipv4 source-ip-address
    100.10.1.0 255.255.255.0
RS G8000(config)# access-control list 4 egress-port 3
RS G8000(config)# access-control list 4 action deny
```

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.

The G8000 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).

VMAPs are configured using the following ISCLI configuration command path:

```
RS G8000(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
  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
```

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

- For a regular VLAN, use config-vlan mode:

```
RS G8000(config)# vlan <VLAN ID>
RS G8000(config-vlan)# [no] vmap <VMAP ID> [serverports|
    non-serverports]
```

- For a VM group (see “VM Group Types” on page 188), use the global configuration mode:

```
RS G8000(config)# [no] virt vmgroup <ID> vmap <VMAP ID>
    [serverports|non-serverports]
```

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 `serverports` or `non-serverports` parameter is specified, the action to add or remove the VMAP is applied for either the switch server ports (`serverports`) or uplink ports (`non-serverports`). If omitted, the operation will be applied to all ports in the associated VLAN or VM group.

Using Storm Control Filters

The G8000 provides filters that can limit the number of the following packet types transmitted by switch ports:

- Broadcast packets
- Multicast packets
- Unknown unicast packets (destination lookup failure)

Broadcast Storms

Excessive transmission of broadcast or multicast traffic can result in a broadcast storm. A broadcast storm can overwhelm your network with constant broadcast or multicast traffic, and degrade network performance. Common symptoms of a broadcast storm are slow network response times and network operations timing out.

Unicast packets whose destination MAC address is not in the Forwarding Database are *unknown unicasts*. When an unknown unicast is encountered, the switch handles it like a broadcast packet and floods it to all other ports in the VLAN (broadcast domain). A high rate of unknown unicast traffic can have the same negative effects as a broadcast storm.

Configuring Storm Control

Configure broadcast filters on each port that requires broadcast storm control. Set a threshold that defines the total number of broadcast packets transmitted (0-2097151), in Megabits per second. When the threshold is reached, no more packets of the specified type are transmitted.

To filter broadcast packets on a port, use the following commands:

```
RS G8000(config)# interface port 1
RS G8000(config-if)# broadcast-threshold <packet rate>
```

To filter multicast packets on a port, use the following commands:

```
RS G8000(config-if)# multicast-threshold <packet rate>
```

To filter unknown unicast packets on a port, use the following commands:

```
RS G8000(config-if)# dest-lookup-threshold <packet rate>
RS G8000(config-if)# exit
```

Part 3: Switch Basics

This section discusses basic switching functions:

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

CHAPTER 7
VLANs

This chapter describes network design and topology considerations for using Virtual Local Area Networks (VLANs). VLANs commonly are 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:

- "VLANs and Port VLAN ID Numbers" on page 98
- "VLAN Tagging" on page 100
- "VLAN Topologies and Design Considerations" on page 104
  This section discusses how you can connect users and segments to a host that supports many logical segments or subnets by using the flexibility of the multiple VLAN system.
- "Protocol-Based VLANs" on page 108
- "Private VLANs" on page 112

Note – VLANs can be configured from the Command Line Interface (see "VLAN Configuration" as well as "Port Configuration" in the *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 RackSwitch G8000 (G8000) 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.

VLANs and Port VLAN ID Numbers

VLAN Numbers

The G8000 supports up to 1024 VLANs per switch. Even though the maximum number of VLANs supported at any given time is 1024, each can be identified with any number between 1 and 4094. VLAN 1 is the default VLAN for the data ports.

Use the following command to view VLAN information:

```
RS G8000)# show vlan

VLAN            Name            Status            Ports
----  ------------------------  ------  -------------------------
1     Default VLAN              ena       1-48, XGE1-XGE4
2     VLAN 2                    dis       empty
```

PVID 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 command to view PVIDs:

```
RS G8000# show interface information

Port Tag    Type     RMON Lrn Fld PVID      NAME            VLAN(s)
---- --- ---------- ---- --- --- ----- ------------ -------------------
1     n  External    d    e   e     1                 1
2     n  External    d    e   e     1                 1
3     n  External    d    e   e     1                 1
4     n  External    d    e   e     1                 1
5     n  External    d    e   e     1                 1
6     n  External    d    e   e     1                 1
...
* = PVID is tagged.
```

Use the following command to set the port PVID:

```
RS G8000(config)# interface port <port number>
RS G8000(config-if)# pvid <PVID number>
```

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” on page 100).

VLAN Tagging

BLADEOS 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.

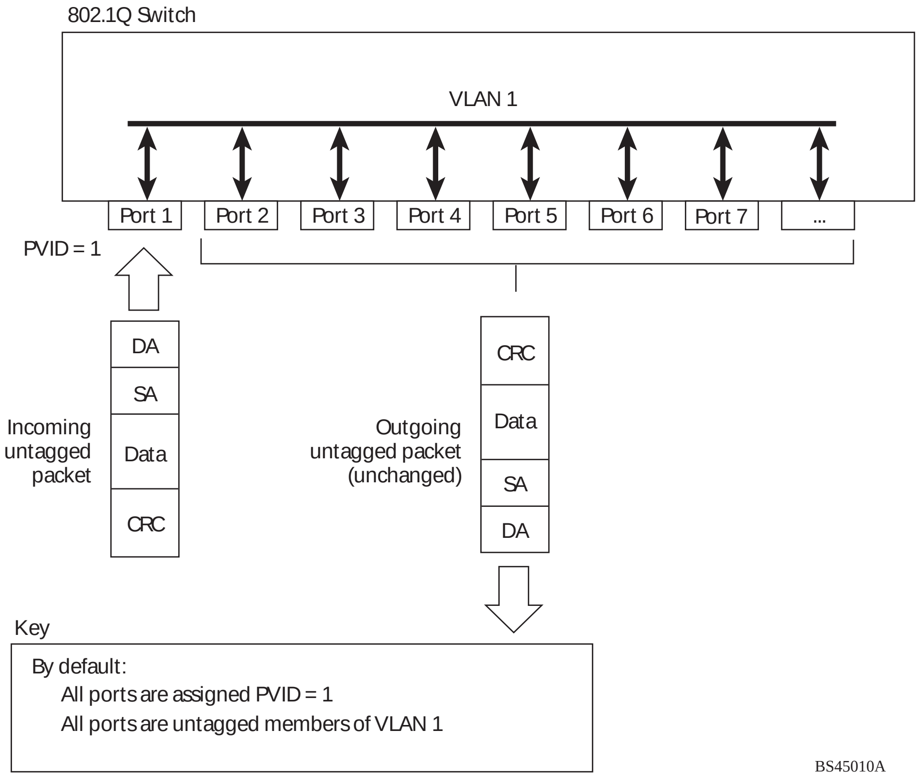
Since tagging fundamentally changes the format of frames transmitted on a tagged port, you must carefully plan network designs to prevent tagged frames from being transmitted to devices that do not support 802.1Q VLAN tags, or devices where tagging is not enabled.

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).

Note – If a 802.1Q tagged frame is received by a port that has VLAN-tagging disabled and the port VLAN ID (PVID) is different than the VLAN ID of the packet, then the frame is dropped at the ingress port.

Figure 2 Default VLAN settings



By default:
All ports are assigned PVID = 1
All ports are untagged members of VLAN 1

Note – The port numbers specified in these illustrations may not directly correspond to the physical port configuration of your switch model.

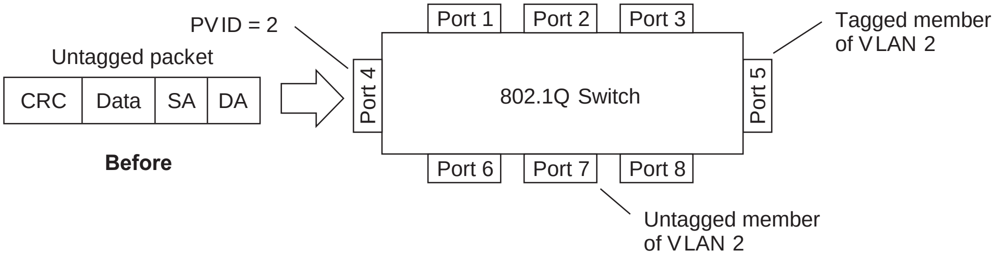
When a VLAN is configured, ports are added as members of the VLAN, and the ports are defined as either *tagged* or *untagged* (see Figure 3 through Figure 6).

The default configuration settings for the G8000 has 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.

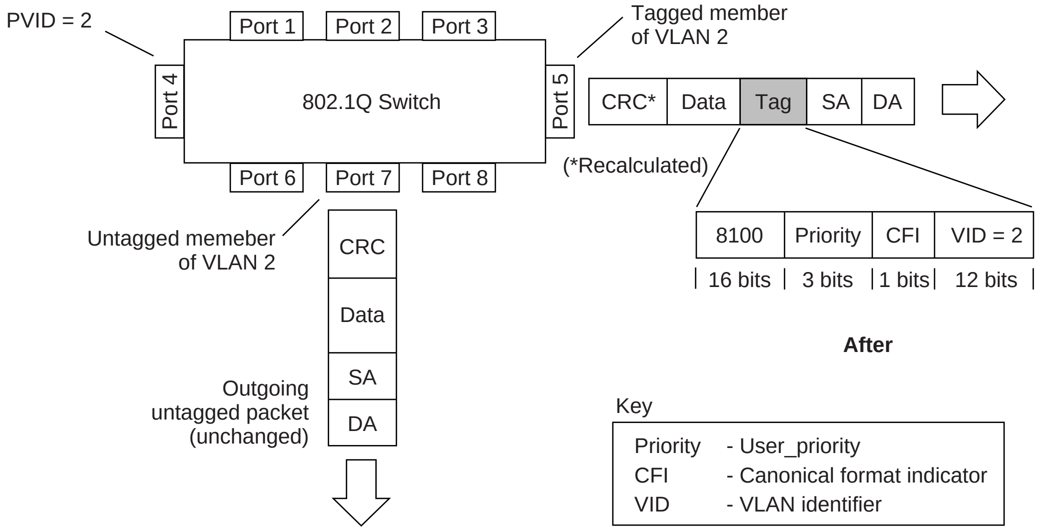
Note – The port assignments in the following figures are not meant to match the G8000.

Figure 3 Port-based VLAN assignment



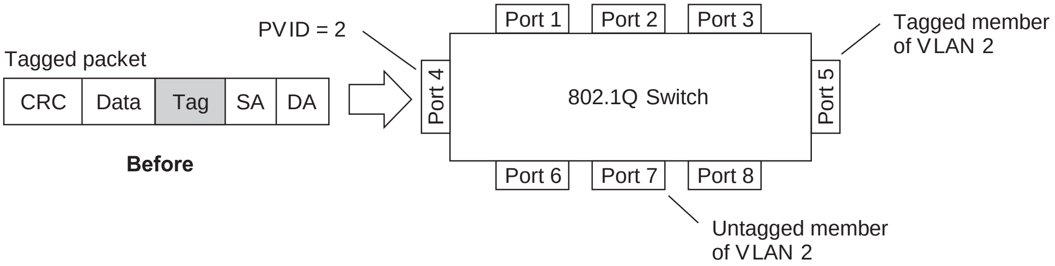
As shown in Figure 4, the untagged packet is marked (tagged) as it leaves the switch through port 5, which is configured as a tagged member of VLAN 2. The untagged packet remains unchanged as it leaves the switch through port 7, which is configured as an untagged member of VLAN 2.

Figure 4 802.1Q tagging (after port-based VLAN assignment)



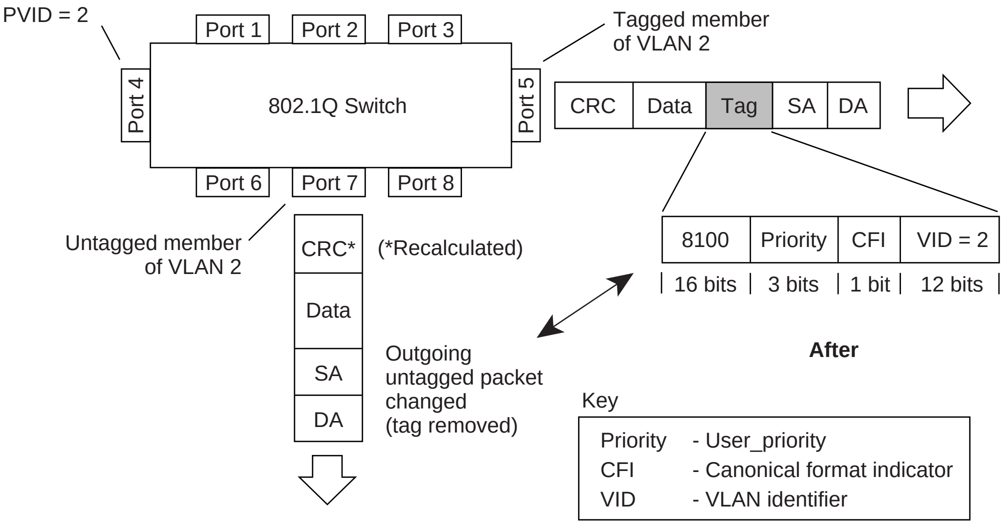
In Figure 5, tagged incoming packets are assigned directly to VLAN 2 because of the tag assignment in the packet. Port 5 is configured as a *tagged* member of VLAN 2, and port 7 is configured as an *untagged* member of VLAN 2.

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)



VLAN Topologies and Design Considerations

- By default, the G8000 software is configured so that tagging is disabled on all ports.
- By default, the G8000 software is configured so that all data ports are members of VLAN 1.
- 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

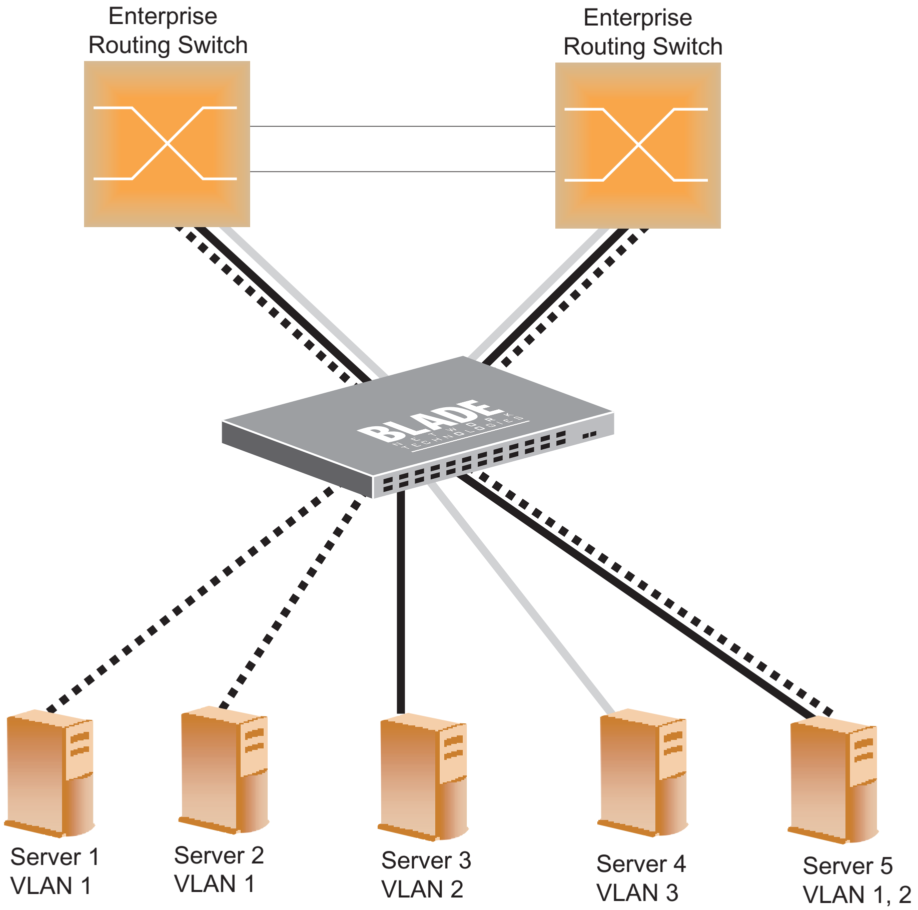
VLANs operate according to specific configuration rules. When creating VLANs, consider the following rules that determine how the configured VLAN reacts in any network topology:

- All ports involved in trunking and port mirroring must have the same VLAN configuration. If a port is on a trunk with a mirroring port, the VLAN configuration cannot be changed. For more information trunk groups, see "Ports and Trunking" on page 115.
- All ports that are involved in port mirroring must have memberships in the same VLANs. 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 371.

Multiple VLANs with Tagging Adapters

Figure 7 illustrates a network topology described in Note – and the configuration example on page page 107.

Figure 7 Multiple VLANs with VLAN-Tagged Gigabit Adapters



The features of this VLAN are described below:

Table 7-1 Multiple VLANs Example

Component	Description
G8000 switch	This switch is configured with three VLANs that represent three different IP subnets. Five ports are connected downstream to servers. Two ports are connected upstream to routing switches. Uplink ports are members of all three VLANs, with VLAN tagging enabled.
Server 1	This server is a member of VLAN 1 and has presence in only one IP subnet. The associated switch port is only a member of VLAN 1, so tagging is disabled.
Server 2	This server is a member of VLAN 1 and has presence in only one IP subnet. The associated switch port is only a member of VLAN 1, so tagging is disabled.
Server 3	This server belongs to VLAN 2, and it is logically in the same IP subnet as Server 5. The associated switch port has tagging disabled.
Server 4	A member of VLAN 3, this server can communicate only with other servers via a router. The associated switch port has tagging disabled.
Server 5	A member of VLAN 1 and VLAN 2, this server can communicate only with Server 1, Server 2, and Server 3. The associated switch port has tagging enabled.
Enterprise Routing switches	These switches must have all three VLANs (VLAN 1, 2, 3) configured. They can communicate with Server 1, Server 2, and Server 5 via VLAN 1. They can communicate with Server 3 and Server 5 via VLAN 2. They can communicate with Server 4 via VLAN 3. Tagging on switch ports is enabled.

Note – VLAN tagging is required only on ports that are connected to other switches or on ports that connect to tag-capable end-stations, such as servers with VLAN-tagging adapters.

VLAN Configuration Example

Use the following procedure to configure the example network shown in Figure 7.

1. Enable VLAN tagging on server ports that support multiple VLANs.

```
RS G8000(config)# interface port 5
RS G8000(config-if)# tagging
RS G8000(config-if)# exit
```

2. Enable tagging on uplink ports that support multiple VLANs.

```
RS G8000(config)# interface port 19
RS G8000(config-if)# tagging
RS G8000(config-if)# exit
RS G8000(config)# interface port 20
RS G8000(config-if)# tagging
RS G8000(config-if)# exit
```

3. Configure the VLANs and their member ports.

```
RS G8000(config)# vlan 2
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 3
RS G8000(config-vlan)# member 5
RS G8000(config-vlan)# member 19
RS G8000(config-vlan)# member 20
RS G8000(config-vlan)# exit
RS G8000(config)# vlan 3
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 4,19,20
RS G8000(config-vlan)# exit
```

By default, all ports are members of VLAN 1, so configure only those ports that belong to other VLANs.

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 1 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 1 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" on page 100.

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 list command (`protocol-vlan` *<x>* `tag-pvlan`) 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 16 VLAN protocols.
- The G8000 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 for ports.

```
RS G8000(config)# interface port 1, 2
RS G8000(config-if)# tagging
RS G8000(config-if)# exit
```

2. Create a VLAN and define the protocol type(s) supported by the VLAN.

```
RS G8000(config)# vlan 2
RS G8000(config-vlan)# enable
Current status: disabled
New status:     enabled
RS G8000(config-vlan)# protocol-vlan 1 frame-type ether2 0800
```

3. Configure the priority value for the protocol.

```
RS G8000(config-vlan)# protocol-vlan 1 priority 2
```

4. Add member ports for this PVLAN.

```
RS G8000(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), you will be asked to confirm changing the PVID to the current VLAN, as shown in the example.

---

5. Enable the PVLAN.

```
RS G8000(config-vlan)# protocol-vlan 1 enable
RS G8000(config-vlan)# exit
```

6. Verify PVLAN operation.

```
RS G8000(config)# show vlan

VLAN            Name            Status            Ports
----  ------------------------  ------  -------------------------
1     Default VLAN              ena      1-48, XGE1-XGE4
2     VLAN 2                    ena      1 2

PVLAN  Protocol  FrameType   EtherType  Priority  Status      Ports
-----  --------  ----------  ---------  --------  -------  -----------
2      1         Ether2      0800       2         enabled  1 2

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.

Use Private VLANs to partition a VLAN domain into sub-domains. Each sub-domain is comprised of one primary VLAN and one or more secondary VLANs, as follows:

- Primary VLAN—carries unidirectional traffic downstream from promiscuous ports. Each Private VLAN configuration 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 and the gateway. Each Private VLAN configuration 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 and the gateway. Each Private VLAN configuration 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). Each promiscuous port can belong to only one Private VLAN.
- 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:

- The default VLAN 1 cannot be a Private VLAN.
- IGMP Snooping must be disabled on isolated VLANs.
- Each secondary port’s (isolated port and community ports) PVID must match its corresponding secondary VLAN ID.
- Ports within a secondary VLAN cannot be members of other VLANs.
- All VLANs that comprise the Private VLAN must belong to the same Spanning Tree Group.

Configuration Example

Follow this procedure to configure a Private VLAN.

1. Select a VLAN and define the Private VLAN type as primary.

```
RS G8000(config)# vlan 100
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 2
RS G8000(config-vlan)# private-vlan type primary
RS G8000(config-vlan)# private-vlan enable
RS G8000(config-vlan)# exit
```

2. Configure a secondary VLAN and map it to the primary VLAN.

```
RS G8000(config)# vlan 110
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 3
RS G8000(config-vlan)# member 4
RS G8000(config-vlan)# private-vlan type isolated
RS G8000(config-vlan)# private-vlan map 100
RS G8000(config-vlan)# private-vlan enable
RS G8000(config-vlan)# exit
```

3. Verify the configuration.

```
RS G8000(config)# show private-vlan

Private-VLAN    Type      Mapped-To   Status          Ports
------------  --------- ---------- ----------  -----------------
100             primary     110        ena          2
110             isolated    100        ena          3-4
```

CHAPTER 8

Ports and Trunking

Trunk groups can provide super-bandwidth, multi-link connections between the RackSwitch G8000 (G8000) and other trunk-capable devices. A trunk group 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 trunking multiple ports together:

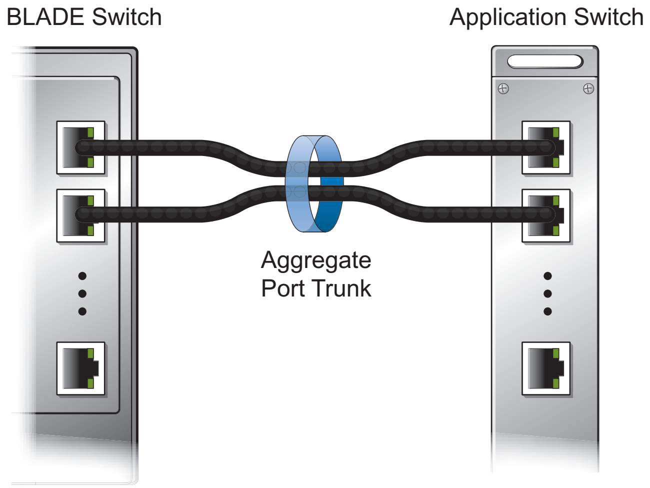
- "Trunking Overview" on this page
- "Port Trunking Example" on page 119
- "Configurable Trunk Hash Algorithm" on page 120
- "Link Aggregation Control Protocol" on page 122

Trunking Overview

When using port trunk groups between two switches, as shown in Figure 8, you can create a virtual link between the switches, operating with combined throughput levels that depends on how many physical ports are included.

Each G8000 supports up to 52 trunk groups in stand-alone (non-stacking) mode, or 64 trunks in stacking mode. Two trunk types are available: static trunk groups (portchannel), and dynamic LACP trunk groups. Each type can contain up to 8 member ports, depending on the port type and availability.

Figure 8 Port Trunk Group



Trunk groups are also useful for connecting a G8000 to third-party devices that support link aggregation, such as Cisco routers and switches with EtherChannel technology (*not* ISL trunking technology) and Sun's Quad Fast Ethernet Adapter. Trunk Group technology is compatible with these devices when they are configured manually.

Trunk traffic is statistically distributed among the ports in a trunk group, based on a variety of configurable options.

Also, since each trunk group is comprised of multiple physical links, the trunk group is inherently fault tolerant. As long as one connection between the switches is available, the trunk remains active and statistical load balancing is maintained whenever a port in a trunk group is lost or returned to service.

Before You Configure Static Trunks

When you create and enable a static trunk, the trunk members (switch ports) take on certain settings necessary for correct operation of the trunking feature.

Before you configure your trunk, you must consider these settings, along with specific configuration rules, as follows:

1. Read the configuration rules provided in the section, "Trunk Group Configuration Rules" on page 118.
2. Determine which switch ports (up to 8) are to become *trunk members* (the specific ports making up the trunk).
3. Ensure that the chosen switch ports are set to `enabled`. Trunk member ports must have the same VLAN and Spanning Tree configuration.
4. Consider how the existing Spanning Tree will react to the new trunk configuration. See Chapter 9, "Spanning Tree Protocols," for Spanning Tree Group configuration guidelines.
5. Consider how existing VLANs will be affected by the addition of a trunk.

Trunk Group Configuration Rules

The trunking feature operates according to specific configuration rules. When creating trunks, consider the following rules that determine how a trunk group reacts in any network topology:

- All trunks must originate from one device, and lead to one destination device.
- Any physical switch port can belong to only one trunk group.
- Trunking from third-party devices must comply with Cisco® EtherChannel® technology.
- All trunk member ports must be assigned to the same VLAN configuration before the trunk can be enabled.
- When an active port is configured in a trunk, the port becomes a *trunk member* when you enable the trunk. The Spanning Tree parameters for the port then change to reflect the new trunk settings.
- All trunk 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 trunk ports can belong to multiple STGs.
- If you change the Spanning Tree participation of any trunk member to `enabled` or `disabled`, the Spanning Tree participation of all trunk members changes similarly.
- When a trunk is enabled, the trunk's Spanning Tree participation setting takes precedence over that of any trunk member.
- You cannot configure a trunk member as a monitor port in a port-mirroring configuration.
- Trunks cannot be monitored by a monitor port; however, trunk members can be monitored.
- All ports in static trunks must be have the same link configuration (speed, duplex, flow control).

Port Trunking Example

In the example below, three ports are trunked between two switches.

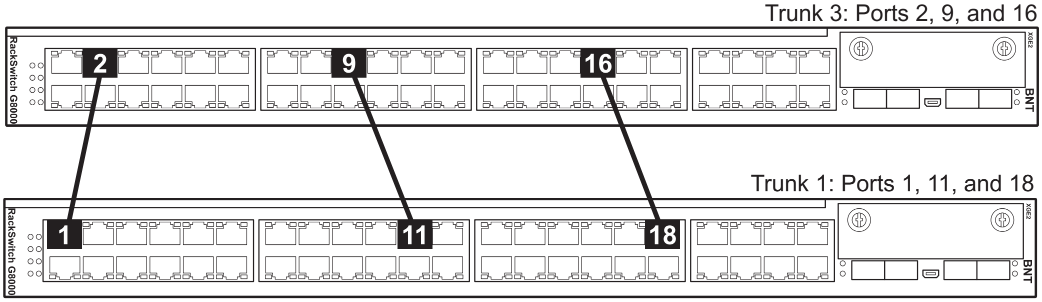


Figure 9 Port Trunk Group Configuration Example

Prior to configuring each switch in the above example, you must connect to the appropriate switches as the administrator.

Note – For details about accessing and using any of the commands described in this example, see the *RackSwitch G8000 ISCLI Reference*.

---

1. Follow these steps on the G8000:

   a. Define a trunk group.

   ```
   RS G8000(config)# portchannel 3 port 2,9,16
   RS G8000(config)# portchannel 3 enable
   ```

   b. Verify the configuration.

   ```
   # show portchannel information
   ```

   Examine the resulting information. If any settings are incorrect, make appropriate changes.

2. Repeat the process on the other switch.

```
RS G8000(config)# portchannel 1 port 1,11,18
RS G8000(config)# portchannel 1 enable
```

3. Connect the switch ports that will be members in the trunk group.

   Trunk group 3 (on the G8000) is now connected to trunk group 1 (on the other switch).

Note – In this example, two G8000 switches 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), trunk groups on the third-party device should be configured manually. Connection problems could arise when using automatic trunk group negotiation on the third-party device.

---

4. Examine the trunking information on each switch.

```
# show portchannel information
PortChannel 3: Enabled
port state:
  2: STG  1 forwarding
  9: STG  1 forwarding
 16: STG  1 forwarding
```

Information about each port in each configured trunk group is displayed. Make sure that trunk groups 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 trunk group.
- Up to 8 ports can belong to the same trunk group.
- All ports in static trunks must be have the same link configuration (speed, duplex, flow control).
- Trunking from third-party devices must comply with Cisco® EtherChannel® technology.

Configurable Trunk Hash Algorithm

Traffic in a trunk group 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 trunk port the frame will use.

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 G8000 supports the following hashing options, which can be used in any combination:

- Frame MAC and IP information. One of the following combinations is required:
  - Source MAC address (`smac`)

    ```
    RS G8000(config)# portchannel hash source-mac-address
    ```

  - Destination MAC address (`dmac`)

    ```
    RS G8000(config)# portchannel hash destination-mac-address
    ```

  - Both source and destination MAC address

    ```
    RS G8000(config)# portchannel hash source-destination-mac
    ```

  - IPv4/IPv6 source IP address (`sip`)

    ```
    RS G8000(config)# portchannel hash source-ip-address
    ```

  - IPv4/IPv6 destination IP address (`dip`)

    ```
    RS G8000(config)# portchannel hash destination-ip-address
    ```

  - Both source and destination IPv4/IPv6 address (enabled by default)

    ```
    RS G8000(config)# portchannel hash source-destination-ip
    ```

- Ingress port number (disabled by default)

  ```
  RS G8000(config)# portchannel hash ingress
  ```

- Layer 4 port information (disabled by default)

  ```
  RS G8000(config)# portchannel hash 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).

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 trunk group or Link Aggregation group) with any device that supports the standard. Please refer to IEEE 802.3ad-2002 for a full description of the standard.

The 802.3ad standard 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 trunk group fails, traffic is reassigned dynamically to the remaining link(s) of the dynamic trunk group.

Note – LACP implementation in the BLADEOS 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 switch port that participates in the same LACP trunk group must have the same *admin key* value. The Admin key is *local significant*, which means the partner switch does not need to use the same Admin key value.

For example, consider two switches, an Actor (the G8000) and a Partner (another switch), as shown in Table 11.

Table 11 Actor vs. Partner LACP configuration

Actor Switch	Partner Switch 1
Port 7 (admin key = 100)	Port 1 (admin key = 50)
Port 8 (admin key = 100)	Port 2 (admin key = 50)

In the configuration shown in Table 11, Actor switch port 7 and port 8 aggregate to form an LACP trunk group with Partner switch port 1 and port 2.

LACP automatically determines which member links can be aggregated and then aggregates them. It provides for the controlled addition and removal of physical links for the link aggregation.

Each port on the switch can have one of the following LACP modes.

- off (default)
  The user can configure this port in to a regular static trunk group.
- active
  The port is capable of forming an LACP trunk. This port sends LACPDU packets to partner system ports.
- passive
  The port is capable of forming an LACP trunk. This port only responds to the LACPDU packets sent from an 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 trunk group 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 trunk group is lost.

When the system is initialized, all ports by default are in LACP *off* mode and are assigned unique *admin keys*. To make a group of ports aggregatable, 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 trunk-forming stage.

Use the following command to check whether the ports are trunked:

```
RS G8000 # show lacp information
```

Note – If you configure LACP on ports with 802.1X network access control, make sure the ports on both sides of the connection are properly configured for both LACP and 802.1X.

LACP Configuration Guidelines

Consider the following guidelines when you configure LACP trunks:

- The range of potential LACP trunk IDs is 53-104.
- When an LACP trunk forms, the trunk ID is determined by the lowest port number in the trunk. For example, if the lowest port number is 1, then the LACP trunk ID is 53.
- Each port that is configured to participate in LACP must be set to full duplex.

Configuring LACP

Use the following procedure to configure LACP for port 7 and port 8 to participate in link aggregation.

1. Configure port parameters. All ports that participate in the LACP trunk group must have the same settings, including VLAN membership.
2. Select the port rant and define the admin key. Only ports with the same admin key can form an LACP trunk group.

```
RS G8000(config)# interface port 7-8
RS G8000(config-if)# lacp key 100
```

3. Set the LACP mode.

```
RS G8000(config-if)# lacp mode active
RS G8000(config-if)# exit
```

CHAPTER 9

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 RackSwitch G8000 (G8000) uses only the most efficient network path.

This chapter covers the following topics:

Spanning Tree Protocol Modes

BLADEOS 6.5 supports the following STP modes:

- Spanning Tree Protocol/Per-VLAN Spanning Tree Plus (STP/PVST+)

  STP as defined in IEEE 802.1D (1998) 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.

  BLADEOS STP/PVST+ supports multiple instances of Spanning Tree, allowing one Spanning Tree Group (STG) per VLAN and is compatible with Cisco PVST+ mode.

  See "STP/PVST+ Mode" on page 127 for details.

- Rapid Spanning Tree Protocol (RSTP)

  IEEE 802.1D (2004) RSTP mode is an enhanced version of STP. It provides more rapid convergence of the Spanning Tree network path states on STG 1.

  RSTP is the default Spanning Tree mode on the G8000. See "Rapid Spanning Tree Protocol" on page 140 for details.

- Per-VLAN Rapid Spanning Tree (PVRST)

  PVRST mode is based on RSTP to provide rapid Spanning Tree convergence, but allows for multiple STGs, with an STGs on a per-VLAN basis. PVRST mode is compatible with Cisco R-PVST/R-PVST+ mode.

  See "Per-VLAN Rapid Spanning Tree Groups" on page 142 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.

  MSTP is supported in stand-alone (non-stacking) mode only.

  See "Multiple Spanning Tree Protocol" on page 143 for details.

Depending on your preferred STG configurations:

Global STP Control

Note – By default, the Spanning Tree feature is globally enabled on the switch, and is set for RSTP mode. Spanning Tree (and thus any currently configured STP mode) can be globally disabled using the following command: If STP is globally disabled, the switch will use STP/PVST+ mode for

```
RS G8000(config)# spanning-tree mode disable
```

internal controls, but will disable Spanning Tree on all user-configurable STGs.

Spanning Tree can be re-enabled by specifying the STP mode:

```
RS G8000(config)# spanning-tree mode {pvst|rstp|pvrst|mst}
```

where the command options represent the following modes:

- `pvst`:STP/PVST+ mode
- `rstp`:RSTP mode
- `pvrst`:PVRST mode
- `mst`:MSTP mode

STP/PVST+ Mode

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.

BLADEOS STP/PVST+ mode implements IEEE 802.1D (1998) Spanning Tree Protocol (STP) with enhancements that allow each VLAN to be assigned to one of 128 available STGs. STP/PVST+ uses IEEE 802.1Q for tagging STP data on a per-VLAN basis, and is compatible with Cisco PVST+ mode. For Cisco R-PVST/R-PVST+ compatibility, see "Per-VLAN Rapid Spanning Tree Groups" on page 142).

The relationship between ports, trunk groups, VLANs, and Spanning Trees is shown in Table 12.

Table 12 Ports, Trunk Groups, and VLANs

Switch Element	Belongs To
Port	Trunk group, or one or more VLANs
Trunk group	One or more VLANs
VLAN (non-default)	One VLAN per STG, or in enhanced modes: ■ RSTP: All VLANs are in STG 1 ■ PVRST: One VLAN per STG ■ MSTP: Multiple VLANs per STG

Port States

STP/PVRST+ mode employs a sequence of port states in the process: Listening, Learning, and Forwarding or Blocking. This process can result in inherent delays for resolving network paths.

To mitigate delays, you can use Port Fast Forwarding (see "Port Fast Forwarding" on page 130) to permit a port that participates in STP/PVST+ to bypass the Listening and Learning states and enter directly into the Forwarding state. While in the Forwarding state, the port listens to the BPDUs to learn if there is a loop and, if dictated by normal STG behavior (following priorities, and so on), the port transitions into the Blocking state.

This feature permits the G8000 to interoperate well within Rapid Spanning Tree networks.

Bridge Protocol Data Units

Bridge Protocol Data Units Overview

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 received BPDU is better than its own BPDU, 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.

Determining the Path for Forwarding BPDUs

When determining which port to use for forwarding and which port to block, the G8000 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:

```
RS G8000(config)# spanning-tree stp <x> 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 command to configure the port priority:

```
RS G8000(config)# spanning-tree stp <STG> priority <priority value>
```

where *priority value* is a number from 0 to 255, in increments of 4 (such as 0, 4, 8, 12, and so on). If the specified priority value is not evenly divisible by 4, the value will be automatically rounded down to the nearest valid increment whenever manually changed in the configuration, or whenever a configuration file from a release prior to BLADEOS 6.5 is loaded.

Note – For RSTP, MSTP, and PVRST modes, port priority must be specified in increments of 16.

Port Path Cost

The port path cost assigns lower values to high-bandwidth ports, such as 10 Gigabit Ethernet, to encourage their use. 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 trunk speed.

Use the following command to modify the port path cost:

```
RS G8000(config)# interface port <port number>
RS G8000(config-if)# spanning-tree stp <STG> path-cost <path cost value>
RS G8000(config-if)# exit
```

The port path cost varies, depending on Spanning Tree mode, as follows:

- STP/PVST+: 1-65535
- RSTP/PVRST/MSTP: 1-200000000 (0 = automatic path cost)

Fast Uplink Convergence

Fast Uplink Convergence enables the G8000 to quickly recover from the failure of the primary link or trunk group in a Layer 2 network using STP/PVST+ mode. Normal recovery can take as long as 50 seconds, while the backup link transitions from Blocking to Listening to Learning and then Forwarding states. With Fast Uplink Convergence enabled, the G8000 immediately places the secondary path into Forwarding state, and multicasts the addresses in the forwarding database (FDB) and ARP table over the secondary link so that upstream switches can learn the new path.

Note – In order for Fast Uplink Convergence to be functional, the switch must be running in STP/PVST mode and must be linked to switches running STP, PVST, or PVST+.

Fast Uplink Configuration Guidelines

When you enable Fast Uplink Convergence, BLADEOS automatically makes the following configuration changes:

- The bridge priority is set to 65535 so that it does not become the root switch.
- The cost of all ports is increased by 3000, across all VLANs and STGs. This ensures that traffic never flows through the G8000 to get to another switch unless there is no other path.

These changes are reversed if the feature is disabled.

Configuring Fast Uplink Convergence

Use the following command to enable Fast Uplink Convergence on ports.

```
RS G8000(config)# spanning-tree uplinkfast
```

Port Fast Forwarding

Port Fast Forwarding permits a port in STP/PVST+ mode to bypass the Listening and Learning states and enter directly into the Forwarding state. While in the Forwarding state, the port listens to the BPDUs to learn if there is a loop and, if dictated by normal STG behavior (following priorities, etc.), the port transitions into the Blocking state. This feature permits the G8000 to interoperate well within Rapid Spanning Tree (RSTP) networks.

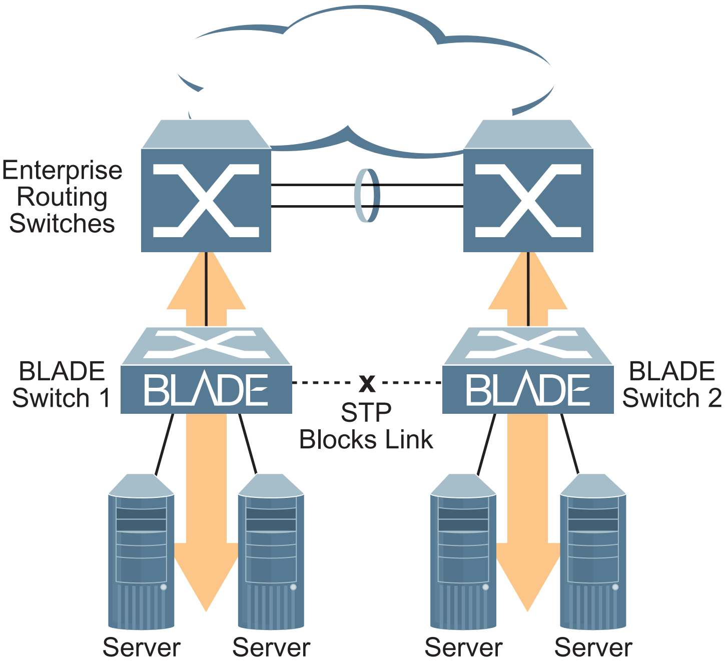
Use the following commands to configure Port Fast Forwarding for a specific STG on a selected port:

```
RS G8000(config)# interface port <port number>
RS G8000(config-if)# [no] spanning-tree stp <STG number> fastforward
RS G8000(config-if)# exit
```

Simple STP Configuration

Figure 10 depicts a simple topology using a switch-to-switch link between two G8000 1 and 2.

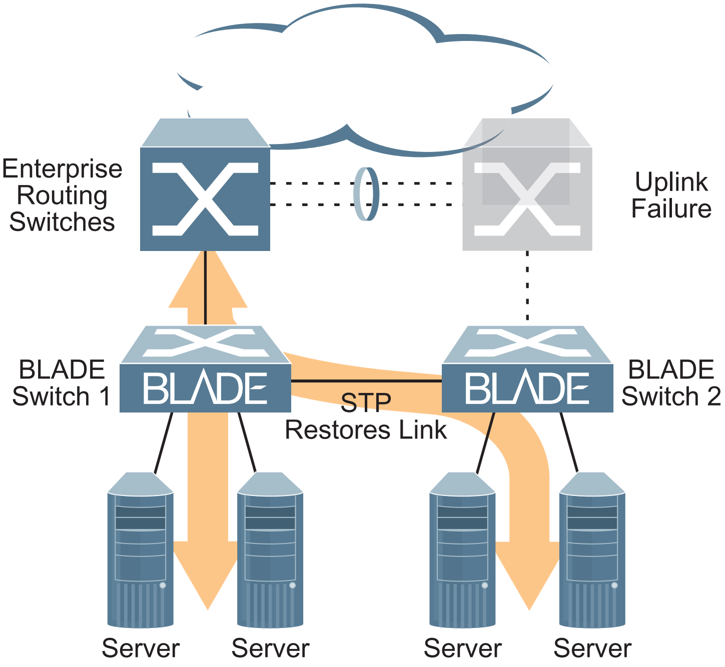
Figure 10 Spanning Tree Blocking a Switch-to-Switch Link



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 G8000 uplinks or the Enterprise switch trunk.

During operation, if one G8000 experiences an uplink failure, STP will activate the BLADE switch-to-switch link so that server traffic on the affected G8000 may pass through to the active uplink on the other G8000, as shown in Figure 11.

Figure 11 Spanning Tree Restoring the Switch-to-Switch Link



In this example, port 10 on each G8000 is used for the switch-to-switch link. To ensure that the G8000 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 G8000.

```
RS G8000(config)# interface port 10
RS G8000(config-if)# spanning-tree stp 1 path-cost 60000
RS G8000(config-if)# exit
```

Per-VLAN Spanning Tree Groups

STP/PVST+ mode supports a maximum of 128 STGs, with each STG acting as an independent, simultaneous instance of STP.

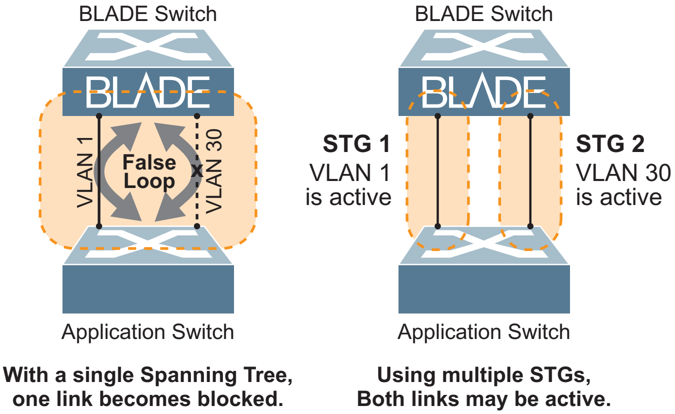
Multiple STGs provide multiple data paths which can be used for load-balancing and redundancy. To enable load balancing between two G8000s using multiple STGs, configure each path with a different VLAN and then assign each VLAN to a separate STG. Since each STG is independent, they each send their own 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 automatically contains any newly configured VLANs 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 12 shows a simple example of why multiple STGs are needed. In the figure, two ports on a G8000 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 G8000, a physical loop is assumed to exist, and one of the VLANs is blocked, impacting connectivity even though no actual loop exists.

Figure 12 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.

STP/PVST+ Defaults and Guidelines

In STP/PVST+ configuration, up to 128 STGs are available on the switch.

STG 1 is the default STG. Although ports 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 and all ports on the switch. Any newly created VLANs will automatically belong to STG 1 until assigned to another STG.

By default, all other STGs (STG 2 through 128) are enabled, though they include no member VLANs or ports. VLANs must be assigned to STGs by the administrator, but 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. Whenever a VLAN is assigned to a new STG, the VLAN is automatically removed from its prior STG.

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.

Adding a VLAN to a Spanning Tree Group

- If no VLANs exist (other than default VLAN 1), see "Creating a VLAN" on page 135 for information creating VLANs and assigning ports to them.
- Otherwise, assign a VLAN to an STG using the following command:

```
RS G8000(config)# spanning-tree stp <STG number> vlan <VLAN number>
```

Note – For proper operation with switches that use Cisco PVST+, it is recommended that you create a separate STG for each VLAN.

Creating a VLAN

- When you create a VLAN, that VLAN automatically belongs to STG 1, the default STG. To place the VLAN in a different STG, follow these steps:
    - ☐ Create the VLAN.
    - ☐ Add the VLAN to an existing STG.

  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 trunked 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 change the port PVID from 3 to 1:

  ```
  "Port 5 is an UNTAGGED port and its PVID changed from 3 to 1.
  ```

- 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 2, 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.

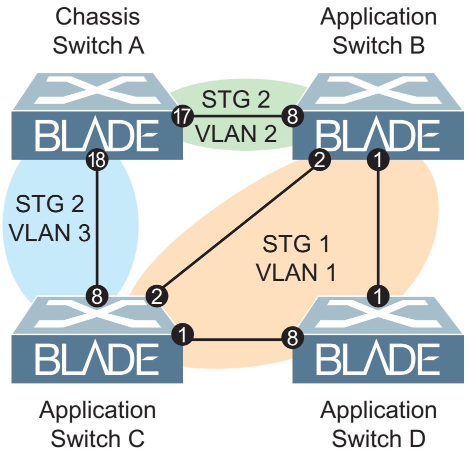
The relationship between port, trunk groups, VLANs, and Spanning Trees is shown in Table 12 on page 127.

Switch-Centric Configuration

STP/PVST+ 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 13, 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 13 Implementing Multiple Spanning Tree Groups



The VLAN participation for each Spanning Tree Group in Figure 13 on page 137 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 2, 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 2. Since Switch C has no additional ports participating in STG 2, 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 13 on page 137.

By default Spanning Trees 2 to 127 are empty, and STG 1 contains all configured VLANs until individual VLANs are explicitly assigned to other STGs.

1. Configure the following on Switch A:

   Add port 17 to VLAN 2, port 18 to VLAN 3, and define STG 2 for VLAN 2 and VLAN 3.

   ```
   RS G8000(config)# vlan 2
   RS G8000(config-vlan)# enable
   RS G8000(config-vlan)# member 17
   RS G8000(config-vlan)# exit
   RS G8000(config)# vlan 3
   RS G8000(config-vlan)# enable
   RS G8000(config-vlan)# member 18
   RS G8000(config-vlan)# exit
   RS G8000(config)# spanning-tree stp 2 vlan 2,3
   ```

   VLAN 2 and VLAN 3 are removed from STG 1.

   Note – In STP/PVST+ mode, each instance of STG is enabled by default.

2. Configure the following on Switch B:

   Add port 8 to VLAN 2 and define STG 2 for VLAN 2.

   ```
   RS G8000(config)# vlan 2
   RS G8000(config-vlan)# enable
   RS G8000(config-vlan)# member 8
   RS G8000(config-vlan)# exit
   RS G8000(config)# spanning-tree stp 2 vlan 2
   ```

   VLAN 2 is automatically removed from STG 1. By default VLAN 1 remains in STG 1.

3. Configure the following on application switch C:

   Add port 8 to VLAN 3 and define STG 2 for VLAN 3.

   ```
   RS G8000(config)# vlan 3
   RS G8000(config-vlan)# enable
   RS G8000(config-vlan)# member 8
   RS G8000(config-vlan)# exit
   RS G8000(config)# spanning-tree stp 2 vlan 3
   ```

   VLAN 3 is automatically removed from STG 1. By default VLAN 1 remains in STG 1.

4. Switch D does not require any special configuration for multiple Spanning Trees. Switch D uses default STG 1 only.

Rapid Spanning Tree Protocol

Note – Rapid Spanning Tree Protocol (RSTP) is enabled by default on the G8000.

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 STP/PVST+ mode STGs 2-128 are not used when the switch is placed in RSTP mode. Although many of the other STP/PVST+ options apply to RSTP as well, there are also new STP parameters to support RSTP, and some values for existing parameters are different.

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.

Port State Changes

The port state controls the forwarding and learning processes of Spanning Tree. In RSTP, the port state has been consolidated to the following: discarding, learning, and forwarding. Table 13 compares the port states between STP/PVST+ mode and RSTP mode.

Table 13 RSTP vs. STP Port states

Operational Status	STP Port State	RSTP Port State
Enabled	Blocking	Discarding
Enabled	Listening	Discarding
Enabled	Learning	Learning
Enabled	Forwarding	Forwarding
Disabled	Disabled	Discarding

Due to Spanning Tree’s sequence of discarding, learning, and forwarding, considerable delays may occur while paths are being resolved. To mitigate delays, ports defined as *edge* ports (“Port Type and Link Type” on page 147) 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 are moved to STG 1.

RSTP Configuration Example

This section provides steps to configure RSTP.

Note – Rapid Spanning Tree is the default Spanning Tree mode on the G8000.

---

1. Configure port and VLAN membership on the switch.
2. Set the Spanning Tree mode to Rapid Spanning Tree.

```
RS G8000(config)# spanning-tree mode rstp
```

3. Configure STP Group 1 parameters.

```
RS G8000(config)# spanning-tree stp 1 enable
RS G8000(config)# spanning-tree stp 1 vlan 2
```

Per-VLAN Rapid Spanning Tree Groups

PVRST is based on IEEE 802.1w Rapid Spanning Tree Protocol (RSTP). Like RSTP, PVRST mode provides rapid Spanning Tree convergence. However, similar to the way standard STP is enhanced by PVST+ (see "Per-VLAN Spanning Tree Groups" on page 133), PVRST is enhanced to allow per-VLAN STGs on the switch.

In PVRST mode, each VLAN may be assigned to one of 128 STGs, with each STG acting as an independent, simultaneous instance of STP. PVRST uses IEEE 802.1Q tagging to differentiate STP BPDUs.

PVRST mode is compatible with Cisco R-PVST/R-PVST+.

Configuring PVRST

This configuration shows how to configure PVRST for VLAN 1 assigned to STG 1, and VLAN2 assigned to STG 2.

1. Set the Spanning Tree mode to PVRST.

```
RS G8000(config)# spanning-tree mode pvrst
```

2. Configure port membership for each VLAN, then define the STGs for each VLAN.

   By default, port 1 is a member of VLAN 1, which automatically assigned to STG 1, so no additional configuration is required for STG 1. However, for STG 2, port 2 if first added to VLAN 2, and then VLAN 2 is assigned to STG 2.

```
RS G8000(config)# vlan 2
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 2
RS G8000(config-vlan)# stg 2
RS G8000(config-vlan)# exit
```

Multiple Spanning Tree Protocol

Note – MSTP is supported in stand-alone (non-stacking) mode only.

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 G8000 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 147) bypass the Discarding and Learning states, and enter directly into the Forwarding state.

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 all VLANs to the CIST. When MSTP is turned off, the switch moves all VLANs from the CIST to STG 1.
- When you enable MSTP, you must configure the Region Name. A default version number of 1 is configured automatically.
- Each bridge in the region must have the same name, version number, and VLAN mapping.

MSTP Configuration Example 1

This section provides steps to configure MSTP on the G8000.

1. Configure port and VLAN membership on the switch.
2. Set the mode to Multiple Spanning Tree, and configure MSTP region parameters.

```
RS G8000(config)# spanning-tree mode mst
RS G8000(config)# spanning-tree mstp name <name>
```

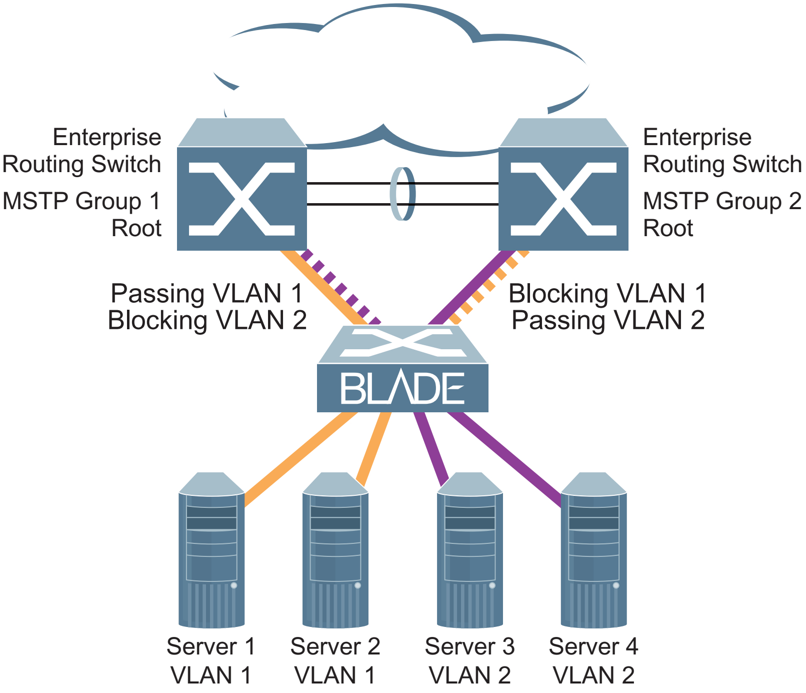
3. Assign VLANs to Spanning Tree Groups.

```
RS G8000(config)# vlan 2
RS G8000(config-vlan)# stg 2
RS G8000(config-vlan)# exit
```

MSTP Configuration Example 2

This configuration shows how to configure MSTP Groups on the switch, as shown in Figure 13.

Figure 14 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.

```
RS G8000(config)# interface port 19
RS G8000(config-if)# tagging
RS G8000(config-if)# exit
RS G8000(config)# interface port 20
RS G8000(config-if)# tagging
RS G8000(config-if)# exit
```

2. Add server ports 1 and 2 to VLAN 1. Add uplink ports 19 and port 20 to VLAN 1.

```
RS G8000(config)# vlan 1
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 1,2,19,20
RS G8000(config-vlan)# stg 1
RS G8000(config-vlan)# exit
```

3. Configure MSTP: Spanning Tree mode, region name, and version.

```
RS G8000(config)# spanning-tree mstp name MyRegion
RS G8000(config)# spanning-tree mode mst
RS G8000(config)# spanning-tree mstp version 100
```

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.

```
RS G8000(config)# vlan 2
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 3,4,5,19,20
RS G8000(config-vlan)# stg 2
RS G8000(config-vlan)# exit
```

Note – Each STG is enabled by default.

Port Type and Link Type

For use in RSTP, MSTP, and PVRST modes, BLADEOS Spanning Tree configuration includes parameters for edge port and link type.

Note – Although edge port and link type parameters are configured with global commands on ports, they only take effect when RSTP, MSTP, or PVRST is turned on.

Edge 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.

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:

```
RS G8000(config)# interface port <port>
RS G8000(config-if)# [no] spanning-tree edge
RS G8000(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:

```
RS G8000(config)# interface port <port>
RS G8000(config-if)# [no] spanning-tree link-type <type>
RS G8000(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 10

Quality of Service

Quality of Service 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 Quality of Service (QoS) level.

The following topics are discussed in this section:

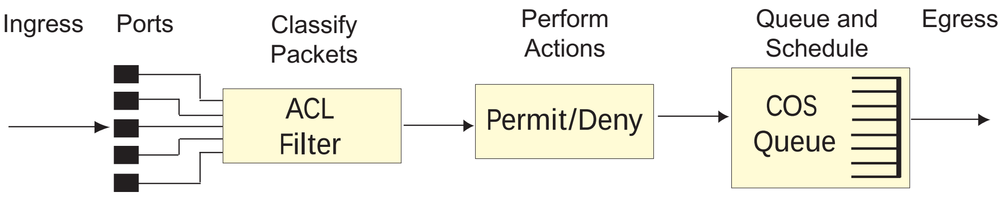
QoS Overview

QoS helps you allocate guaranteed bandwidth to the 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 that cannot tolerate delay, by assigning their 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 15 shows the basic QoS model used by the switch.

Figure 15 QoS Model



The basic QoS model works as follows:

- Classify traffic:
  - Read DSCP value.
  - Read 802.1p priority value.
  - Match ACL filter parameters.
- Perform actions:
  - Define bandwidth and burst parameters
  - Select actions to perform on in-profile and out-of-profile traffic
  - Deny packets
  - Permit 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 COS queues.
  - Schedule transmission based on the COS queue.

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 the conditions that must match for inclusion in the filter, and also the actions that are performed when a match is made.

BLADEOS 6.5 supports up to 512 ACLs.

The G8000 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.
- TCP/UPD: Source port, destination port, TCP flag.
- Packet format

For ACL details, see "Access Control Lists" on page 81.

Summary of ACL Actions

Actions determine how the traffic is treated. The G8000 QoS actions include the following:

- Pass or Drop
- Re-mark 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 G8000 by configuring a QoS meter (if desired) and assigning ACLs to ports. When you add ACLs 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.

Using meters, you set a Committed Rate in Kbps (multiples of 64 Mbps). All traffic within this Committed Rate is In-Profile. Additionally, you 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 switch uses the Differentiated Services (DiffServ) architecture to provide QoS functions. DiffServ is described in IETF RFCs 2474 and 2475.

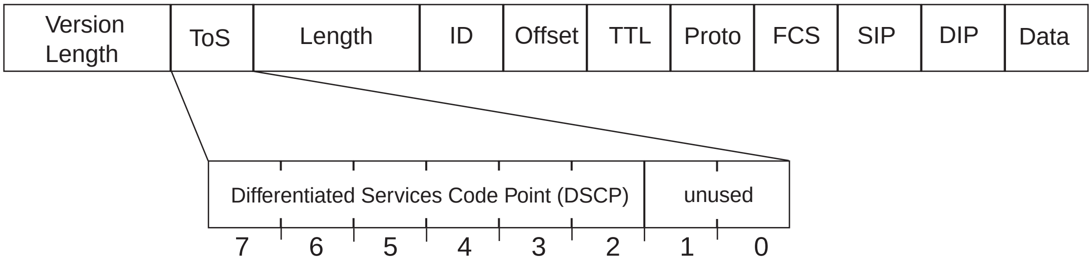
The six most significant bits in the TOS byte of the IP header are defined as DiffServ Code Points (DSCP). Packets are marked with a certain value depending on the type of treatment the packet must receive in the network device. DSCP is a measure of the Quality of Service (QoS) level of the packet.

The switch 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 switch 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.

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 16 Layer 3 IPv4 packet



The switch 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 a Class of Service queue (COSq).

The switch can use the DSCP value to direct traffic prioritization.

With DiffServ, you can establish policies to direct 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.

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 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 14 shows the default service levels provided by the switch, listed from highest to lowest importance:

Table 14 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 switch can use the DSCP value of ingress packets to re-mark the DSCP to a new value, and to set an 802.1p priority value. Use the following command to view the default settings.

```
RS G8000# show qos dscp
Current DSCP Remarking Configuration: OFF

  DSCP    New DSCP  New 802.1p Prio
--------  --------  ---------------
    0         0          0
    1         1          0
    2         2          0
    3         3          0
    4         4          0
    5         5          0
    6         6          0
    7         7          0
    8         8          1
    9         9          0
   10        10          1
...
   54        54          0
   55        55          0
   56        56          7
   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:

```
RS G8000(config)# qos dscp re-marking
```

Then you must enable DSCP re-marking on any port that you wish to perform this function (Interface Port mode).

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. Turn DSCP re-marking on globally, and define the DSCP-DSCP-802.1p mapping. You can use the default mapping.

```
RS G8000(config)# qos dscp re-marking
RS G8000(config)# qos dscp dscp-mapping <DSCP value (0-63)> <new value>
RS G8000(config)# qos dscp dot1p-mapping <DSCP value (0-63)> <802.1p value>
```

2. Enable DSCP re-marking on a port.

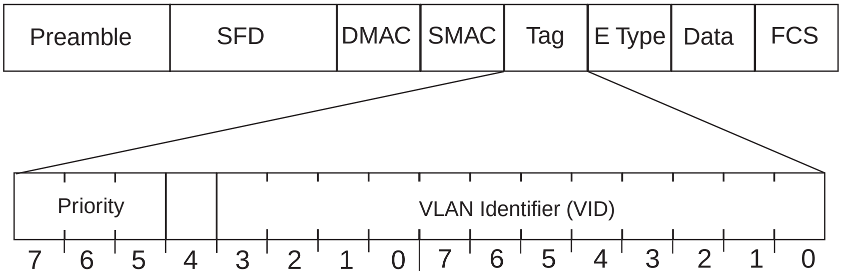
```
RS G8000(config)# interface port 1
RS G8000(config-if)# qos dscp dscp-remarking
```

Using 802.1p Priority to Provide QoS

The G8000 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 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 switch can filter packets based on the 802.1p values.

Figure 17 Layer 2 802.1q/802.1p VLAN tagged packet



Ingress packets receive a priority value, as follows:

- **Tagged packets**—switch reads the 802.1p priority in the VLAN tag.
- **Untagged packets**—switch tags the packet and assigns an 802.1p priority value, based on the port’s default 802.1p 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.

```
RS G8000(config)# interface port 1
RS G8000(config-if)# dot1p <802.1p value (0-7)>
```

Queuing and Scheduling

The G8000 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.

Note – In stacking mode, because one COS queue is reserved for internal use, the number of configurable COS queues is either 1 or 7.

Each COS queue uses Weighted Round Robin (WRR) scheduling, with user configurable weight from 1 to 15. The weight of 0 (zero) indicates strict priority, which might starve the low priority queues.

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:

```
RS G8000(config)# qos transmit-queue mapping <802.1p priority value (0-7)>
      <COS queue (0-7)>
```

To set the COS queue scheduling weight, use the following command.

```
RS G8000(config)# qos transmit-queue weight-cos <COSq number>
      <COSq weight (0-15)>
```

Part 4: Advanced Switching Features

CHAPTER 11

Stacking

This chapter describe how to implement the stacking feature in the RackSwitch G8000. The following concepts are covered:

Stacking Overview

A *stack* is a group of up to six RackSwitch G8000 switches with BLADEOS that work together as a unified system. 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.
- Once the stacking links have been established (see below), 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, Trunks, Trunk 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.

Stacking Requirements

Before BLADEOS switches can form a stack, they must meet the following requirements:

- All switches must be the same model (RackSwitch G8000).
- Each switch must be installed with BLADEOS, version 6.5 or later. The same release version is not required, as the Master switch will push a firmware image to each differing switch which is part of the stack.
- The recommended stacking topology is a bidirectional ring (see Figure 18 on page 172). To achieve this, two 10Gb Ethernet ports on each switch must be reserved for stacking. By default, 10Gb Ethernet ports 51 and 52 (via optional 10GbE modules installed at the back of the switch) are used.
- 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 in order to avoid internal stack reconfiguration.

Stacking Limitations

The G8000 with BLADEOS 6.5 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:

- Active Multi-Path Protocol (AMP)
- BCM rate control
- Border Gateway Protocol (BGP)
- IGMP Relay and IGMPv3
- IPv6
- Link Layer Detection Protocol (LLDP)
- Loopback Interfaces
- MAC address notification
- MSTP
- OSPF and OSPFv3
- Port flood blocking
- Protocol-based VLANs
- Protocol Independent Multicast (PIM)
- RIP
- Router IDs
- Route maps
- SFD
- sFlow port monitoring
- Static MAC address adding
- Static multicast
- Uni-Directional Link Detection (UDLD)
- 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 six switches, interconnected by a stack trunk in a local ring topology (see Figure 18 on page 172). 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 should 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, although uncommon, there are scenarios in which a stack may temporarily have more than one Master switch. Should this occur, 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 Members which cannot access either the Master or Backup are considered *isolated* and will not process network traffic (see “No Backup” on page 169). 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.

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.

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 in order 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, the switches 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.

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 and lose all settings pertaining to their original stacks.

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 setting is deleted or changed using the following commands from the active Master:

```
RS G8000(config)# no stack backup

 -or-

RS G8000(config)# stack backup <csnum 1-6>
```

- 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.

Secondary Backup

When a Backup takes over stack control operations, if any other configured Masters (acting as Member switches) are available within the stack, the Backup will select one as a secondary Backup. The primary Backup automatically reconfigures the secondary Backup and specifies itself (the primary Backup) as the new Backup in case the secondary fails. This prevents the chain of stack control from migrating too far from the original Master and Backup configuration intended by the administrator.

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 `WAITING` state until a Master appears. During this `WAITING` 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

Configuration Overview

This section provides procedures for creating a stack of switches. The high-level procedure is as follows:

- 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.
- Configure an external IP interface on the Master (if external management is desired).
- 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 18 on page 172).
- Avoid disrupting the stack connections unnecessarily while the stack is in operation.
- For enhanced redundancy when creating port trunks, include ports from different stack members in the trunks.
- 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.

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:

```
RS G8000(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:

```
RS G8000(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 below, be reserved for stacking.

```
RS G8000(config)# boot stack vlan 4090
```

4. On each switch, designate the stacking links.

   To create the recommended topology, at least two 10Gb ports on each switch should be dedicated to stacking. By default, 10Gb Ethernet ports 51 and 52 (via optional 10GbE modules installed at the back of the switch) are used. Use the following command to specify the links to be used in the stacking trunk:

```
RS G8000(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.

---

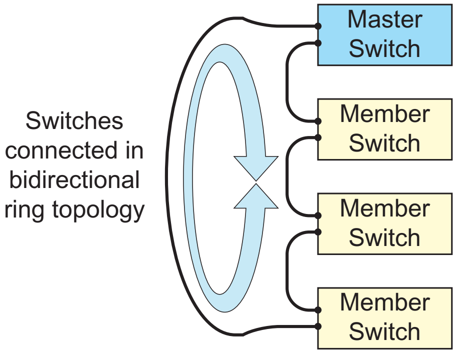
5. On each switch, perform a reboot:

```
RS G8000(config)# reload
```

6. Physically connect the stack trunks.

   To create the recommended topology, attach the two designated stacking links in a bidirectional ring. As shown in Figure 18, 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 18 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 trunks 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.

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.

Configuring an External IPv4 Address for the Stack

In addition to the internal management IP interface assigned to the Master switch by the management system, a standard switch IP interface can be used for connecting to and managing the stack externally. Configure an IP interface with the following:

- Stack IPv4 address and mask
- IPv4 default gateway address
- VLAN number used for external access to the stack (rather than the internal VLAN 4090 used for inter-stack traffic)

One completed, stack management can be performed via Telnet or BBI (if enabled) from any point in the configured VLAN, using the IPv4 address of the configured IP interface.

```
RS G8000(config)# ip gateway <gateway number> address <gateway IPv4 address>
RS G8000(config)# ip gateway <gateway number> enable
RS G8000(config)# interface ip <IP interface number>
RS G8000(config-ip-if) ip address <stack IPv4 address>
RS G8000(config-ip-if) ip netmask <IPv4 subnet mask>
RS G8000(config-ip-if) vlan <VLAN ID>
RS G8000(config-ip-if) enable
RS G8000(config-ip-if) exit
```

In the event that the Master switch fails, if a Backup switch is configured (see "Assigning a Stack Backup Switch" on page 175), the external IP interface for the stack will still be available. Otherwise, the administrator must manage the stack through the internal management IP interface assigned to the Backup switch by the management system.

Locating an External Stack Interface

If the IPv4 address and VLAN of an external IP interface for the stack is unknown, connect to the Master switch using the IPv4 address assigned by the management system, and execute the following command:

```
RS G8000(config)# show interface ip
```

Viewing Stack Connections

To view information about the switches in a stack, execute the following command:

```
RS G8000(config)# show stack switch

Stack name:
Local switch is the master.

Local switch:
  csnum - 1
  MAC - 00:00:00:00:01:00
  Switch Type - 9
  Chassis Type - 99
  Switch Mode (cfg) - Master
  Priority - 225
  Stack MAC - 00:00:00:00:01:1f

Master switch:
  csnum - 1
  MAC - 00:00:00:00:01:00

Backup switch:
  csnum - 2
  MAC - 00:22:00:ad:43:00

Configured Switches:
--------------------------------
csnum        MAC         asnum
--------------------------------
  C1  00:00:00:00:01:00  A1
  C2  00:22:00:ad:43:00  A3
  C3  00:11:00:af:ce:00  A2

Attached Switches in Stack:
-------------------------------------
asnum        MAC        csnum   State
-------------------------------------
  A1  00:00:00:00:01:00  C1  IN_STACK
  A2  00:11:00:af:ce:00  C3  IN_STACK
  A3  00:22:00:ad:43:00  C2  IN_STACK
```

Binding Members to the Stack

You can bind Member switches to a stack `csnum` using either their `asnum` or MAC address :

```
RS G8000(config)# stack switch-number <csnum> mac <MAC address>
    -or-
RS G8000(config)# stack switch-number <csnum> bind <asnum>
```

To remove a Member switch, execute the following command:

```
RS G8000(config)# no stack switch-number <csnum>
```

Assigning a Stack Backup Switch

To define a Member switch as a Backup (optional) which will assume the Master role if the Master switch should fail, execute the following command:

```
RS G8000(config)# stack backup <csnum>
```

Managing a Stack

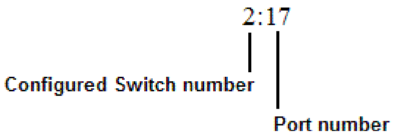
The stack is managed primarily through the Master switch. The Master switch then pushes configuration changes and run-time information to the Member switches.

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 trunk, and use the IP address of any IP interface to access the stack.

Stacking Port Numbers

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.

Note – Do not use VLAN 4090 for any purpose other than internal stacking traffic.

Rebooting Stacked Switches using the ISCLI

The administrator can reboot individual switches in the stack, or the entire stack using the following commands:

```
RS G8000(config)# reload                       (Reboot all switches in the stack)
RS G8000(config)# reload master                (Reboot only the stack Master)
RS G8000(config)# reload switch <csnum list>   (Reboot only the listed 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 15 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 an Existing Stack

Upgrade all stacked switches at the same time. The Master controls the upgrade process. Use the following procedure to perform a software upgrade for a stacked system.

1. Load new software on the Master.

   The Master pushes the new software image to all Members in the stack, as follows:

   - If the new software is loaded into image 1, the Master pushes the software into image 1 on all Members.
   - If loaded into image 2, the Master pushes the software into image 2 on all Members.

   The software push can take several minutes to complete.

2. Verify that the software push is complete. Use either the BBI or the ISCLI:

   - From the BBI, go to Dashboard > Stacking > Push Status and view the Image Push Status Information, or
   - From the ISCLI, use following command to verify the software push:

```
RS G8000(config)# show stack push-status

Image 1 transfer status info:
       Switch 00:16:60:f9:33:00:
              last receive successful
       Switch 00:17:ef:c3:fb:00:
              not received - file not sent or transfer in progress

Image 2 transfer status info:
       Switch 00:16:60:f9:33:00:
              last receive successful
       Switch 00:17:ef:c3:fb:00:
              last receive successful

Boot image transfer status info:
       Switch 00:16:60:f9:33:00:
              last receive successful
       Switch 00:17:ef:c3:fb:00:
              last receive successful

Config file transfer status info:
       Switch 00:16:60:f9:33:00:
              last receive successful
       Switch 00:17:ef:c3:fb:00:
              last receive successful
```

3. Reboot all switches in the stack. Use either the ISCLI or the BBI.

   - From the BBI, select Configure > System > Config/Image Control. Click Reboot Stack.
   - From the ISCLI, use the following command:

   ```
   RS G8000(config)# reload
   ```

4. Once the switches in the stack have rebooted, verify that all of them are using the same version of firmware. Use either the ISCLI or the BBI.

   - From the BBI, open Dashboard > Stacking > Stack Switches and view the Switch Firmware Versions Information from the Attached Switches in Stack.
   - From the ISCLI, use the following command:

   ```
   RS G8000(config)# show stack version
   Switch Firmware Versions:
   ------------------------------------------------------------
   asnum  csnum         MAC           S/W    Version   Serial #
   -----  -----  -----------------  ------  -------  ----------
    A1     C1    00:00:00:00:01:00  image1  0.0.0.0  CH49000000
    A2     C2    00:11:00:af:ce:00  image1  0.0.0.0  CH49000001
    A3           00:22:00:ad:43:00  image1  0.0.0.0  CH49000002
   ```

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 in order 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 from the chassis.

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 164.
2. Place the new switch in its determined place according to the *RackSwitch G8000 Installation Guide*.

3. Connect to the ISCLI of the new switch (not the stack interface)
4. Enable stacking:

```
RS G8000(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:

     ```
     RS G8000(config)# boot stack mode member
     ```

   - If replacing a Master switch:

     ```
     RS G8000(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 (shown below).

   ```
   RS G8000(config)# boot stack vlan 4090
   ```

7. Designate the stacking links.

   It is recommended that you designate the same number of 10Gb ports for stacking as the switch being replaced. By default, 10Gb Ethernet port 51 and port 52 (via optional 10GbE modules installed at the back of the switch) are used. At least one 10Gp port is required. Use the following command to specify the links to be used in the stacking trunk:

   ```
   RS G8000(config)# boot stack higig-trunk <list of port names or aliases>
   ```

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:

    ```
    RS G8000(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 :

```
RS G8000(config)# stack switch-number <csnum> mac <MAC address>
    -or-
RS G8000(config)# stack switch-number <csnum> bind <asnum>
```

3. Apply and save your configuration changes.

Note – If replacing the Master switch, the Master will not assume control from the Backup unless the Backup is rebooted or fails.

ISCLI Stacking Commands

Stacking-related ISCLI commands are listed below. For details on specific commands, see the *RackSwitch G8000 ISCLI Reference*.

- `[no] boot stack enable`
- `boot stack higig-trunk` *<port list>*
- `boot stack mode master|member`
- `boot stack push-image boot-image|image1|image2` *<asnum>*
- `boot stack vlan` *<VLAN>* *<asnum>*`|master|backup|all`
- `default boot stack` *<asnum>*`|master|backup|all`
- `[no] logging log stacking`
- `no stack backup`
- `no stack name`
- `no stack switch-number` *<csnum>*
- `show boot stack` *<asnum>*`|master|backup|all`
- `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>* `bind` *<asnum>*
- **`stack switch-number`** *<csnum>* **`mac`** *<MAC address>*

CHAPTER 12
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 BLADEOS 6.5 on the RackSwitch G8000 (G8000):

- 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 97.

- Port trunking

  A port trunk pools multiple physical switch ports into a single, high-bandwidth logical link to other devices. In addition to aggregating capacity, trunks provides link redundancy.

  For details on this feature, see “Ports and Trunking” on page 115.

- Stacking

  Multiple switches can be aggregated into a single super-switch, combining port capacity while at the same time simplifying their management. BLADEOS 6.5 supports one stack with up to six switches.

  For details on this feature, see “Stacking” on page 163.

- 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 187.

BLADEOS virtualization features provide a highly-flexible framework for allocating and managing switch resources.

CHAPTER 13

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 BLADEOS 6.5 VMready feature supports up to 1024 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 G8000 with VMready also detects the migration of VEs across different hypervisors. As VEs move, the G8000 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 G8000 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. BLADEOS 6.5 supports up to 1024 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).

Defining Server Ports

Before you configure VMready features, you must first define whether ports are connected to servers or are used as uplink ports. Use the following ISCLI configuration command to define a port as a server port:

```
RS G8000(config)# system server-ports port <port alias or number>
```

Ports that are not defined as server ports are automatically considered uplink ports.

VM Group Types

VEs, as well as switch server ports, switch uplink ports, static trunks and LACP trunks, 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.

BLADEOS 6.5 supports up to 32 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 194).

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 trunks, 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 G8000, may require manual configuration when using local VM groups.

Configuring a Local VM Group

Use the following ISCLI configuration commands to assign group properties and membership:

```
RS G8000(config)# virt vmgroup <VM group number> ?
  key <LACP trunk key>                                   (Add LACP trunk to group)
  port <port alias or number>                            (Add port member to group)
  portchannel <trunk group number>                       (Add static trunk to group)
  profile <profile name>                                 (Not used for local groups)
  stg <Spanning Tree group>                              (Add STG to group)
  tag                                                    (Set VLAN tagging on ports)
  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)
```

The following rules apply to the local VM group configuration commands:

- `key`: Add LACP trunks to the group.
- `port`: Add switch server ports or switch uplink ports to the group.
- `portchannel`: Add static port trunks 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 191).
- `stg`: The group may be assigned to a Spanning-Tree group for broadcast loop control (see "Spanning Tree Protocols" on page 125).
- `tag`: Enable VLAN tagging for the VM group. If the VM group contains ports which also exist in other VM groups, tagging should be enabled in both VM groups.
- `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 197).
- `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 200).

  If VMware Tools software is installed in the guest operating system (see VMware documentation for information on installing recommended tools), VEs may also be specified by IPv4 address or VE name. However, if there is more than one possible VE for the input, the switch will display a list of candidates and prompt for a specific MAC address.

  Only VEs currently connected to the switch port (local) or pending connection (pre-provisioned) are permitted in local VM groups.

Use the `no` variant of the commands to remove or disable VM group configuration settings:

```
RS G8000(config)# no virt vmgroup <VM group number> [?]
```

Distributed VM Groups

Distributed VM groups allow configuration profiles to be synchronized between the G8000 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 G8000 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 194 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:

```
RS G8000(config)# virt vmprofile <profile name>
```

Next, the profile must be edited and configured using the following configuration commands:

```
RS G8000(config)# virt vmprofile edit <profile name> ?
  vlan <VLAN number>
  shaping <average bandwidth> <burst size> <peak>
```

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 198.

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 191 for details.

Once a VM profile is available, a distributed VM group may be initialized using the following configuration command:

```
RS G8000(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 using the following ISCLI configuration command:

```
RS G8000(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 189) cannot be converted and must be deleted before it can be used for a distributed VM group.

Assigning Members

VMs, ports, and trunks 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 189), 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 190) 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 server 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:

`BNT_`<*VM profile name*>

Adding 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 `BNT_Default` 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.

Virtualization Management Servers

The G8000 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. BLADEOS 6.5 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 G8000 to associate the vCenter with the switch:

```
RS G8000(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 G8000 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:

```
RS G8000# virt vmware scan                        (Scan the vCenter)
    -or-
RS G8000# 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:

```
RS G8000(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 G8000 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:

```
RS G8000# 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 G8000 may be used as a central point of configuration for VMware virtual switches and port groups using the following ISCLI privileged EXEC commands:

```
RS G8000# virt vmware ?
  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:

```
RS G8000(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 81). 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.

BLADEOS 6.5 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).

VMAPs are configured using the following ISCLI configuration command path:

```
RS G8000(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
  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
```

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

- For regular VLANs, use config-vlan mode:

```
RS G8000(config)# vlan <VLAN ID>
RS G8000(config-vlan)# [no] vmap <VMAP ID> [serverports|
    non-serverports]
```

- For a VM group, use the global configuration mode:

```
RS G8000(config)# [no] virt vmgroup <ID> vmap <VMAP ID>
      [serverports|non-serverports]
```

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 `serverports` or `non-serverports` parameter can be specified to apply the action (to add or remove the VMAP) for either the switch server ports (`serverports`) or switch uplink ports (`non-serverports`). 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 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:

```
RS G8000(config)# virt vmpolicy vmbwidth <VM MAC>|<index>|<UUID>|
      <IPv4 address>|<name> ?
  txrate <committed rate> <burst> [<ACL number>]       (Set the VM to switch rate)
  bwctrl                                               (Enable bandwidth control)
```

Bandwidth allocation can be defined for transmit (TX) trafficonly. 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.

The *committed rate* is specified in multiples of 64 kbps, from 64 to 10,000,000. 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 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 191) 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 G8000. 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 G8000. Shaping is not supported simultaneously on the same ports as vNICs.

VM Policy Bandwidth Control is configured per VE, and can be set independently for transmit traffic. Bandwidth policies are enforced by the G8000. VE traffic that exceeds configured levels is dropped by the switch upon ingress. Setting `txrate` uses ACL resources on the switch.

Bandwidth shaping and bandwidth policies can be used separately or in concert.

VMready Information Displays

The G8000 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 194), 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 195.

Local VE Information

A concise list of local VEs and pre-provisioned VEs is available with the following ISCLI privileged EXEC command:

```
RS G8000# show virt vm

IP Address       VMAC Address         Index Port    VM Group (Profile)
---------------- -----------------    ----- ------- ------------------
*172.16.46.50    00:50:56:4e:62:00    4     3
*172.16.46.10    00:50:56:4f:f2:00    2     4
+172.16.46.51    00:50:56:72:ec:00    1     3
+172.16.46.11    00:50:56:7c:1c:00    3     4
 172.16.46.25    00:50:56:9c:00:00    5     4
 172.16.46.15    00:50:56:9c:21:00    0     4
 172.16.46.35    00:50:56:9c:29:00    6     3
 172.16.46.45    00:50:56:9c:47:00    7     3

 Number of entries: 8
 * indicates VMware ESX Service Console Interface
 + indicates VMware ESX/ESXi VMKernel or Management Interface
```

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 option:

```
RS G8000# 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 203).

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:

```
RS G8000# 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:

```
RS G8000# 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:

```
RS G8000# 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:

```
RS G8000# 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 switch server ports 1 and 2, and one static trunk (previously configured) that includes switch uplink ports 3 and 4.

1. Define the server ports.

```
RS G8000(config)# system server-ports port 1-2
```

2. Enable the VMready feature.

```
RS G8000(config)# virt enable
```

3. Specify the VMware vCenter IPv4 address.

```
RS G8000(config)# virt vmware vmware vcspec 172.16.100.1 bladevm
```

When prompted, enter the user password that the switch must use for access to the vCenter.

4. Create the VM profile.

```
RS G8000(config)# virt vmprofile Finance
RS G8000(config)# virt vmprofile edit Finance vlan 30
RS G8000(config)# virt vmprofile edit Finance shaping 1000 2000 3000
```

5. Define the VM group.

```
RS G8000(config)# virt vmgroup 1 profile Finance
RS G8000(config)# virt vmgroup 1 vm arctic
RS G8000(config)# virt vmgroup 1 vm monster
RS G8000(config)# virt vmgroup 1 vm sierra
RS G8000(config)# virt vmgroup 1 vm 00:50:56:4f:f2:00
RS G8000(config)# virt vmgroup 1 portchannel 1
```

When VMs are added, the server ports on which they appear are automatically added to the VM group. In this example, there is no need to manually add ports 1 and 2.

6. If necessary, enable VLAN tagging for the VM group:

```
RS G8000(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.

---

7. Save the configuration.

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
- IPv6 Host Management
- Routing Information Protocol (RIP)
- Internet Group Management Protocol (IGMP)
- Border Gateway Protocol (BGP)
- Open Shortest Path First (OSPF)
- Protocol Independent Multicast (PIM)

CHAPTER 14

Basic IP Routing

This chapter provides configuration background and examples for using the G8000 to perform IP routing functions. The following topics are addressed in this chapter:

- "IP Routing Benefits" on page 209
- "Routing Between IP Subnets" on page 209
- "Example of Subnet Routing" on page 211
- "ECMP Static Routes" on page 215
- "Dynamic Host Configuration Protocol" on page 217

IP Routing Benefits

The switch 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.

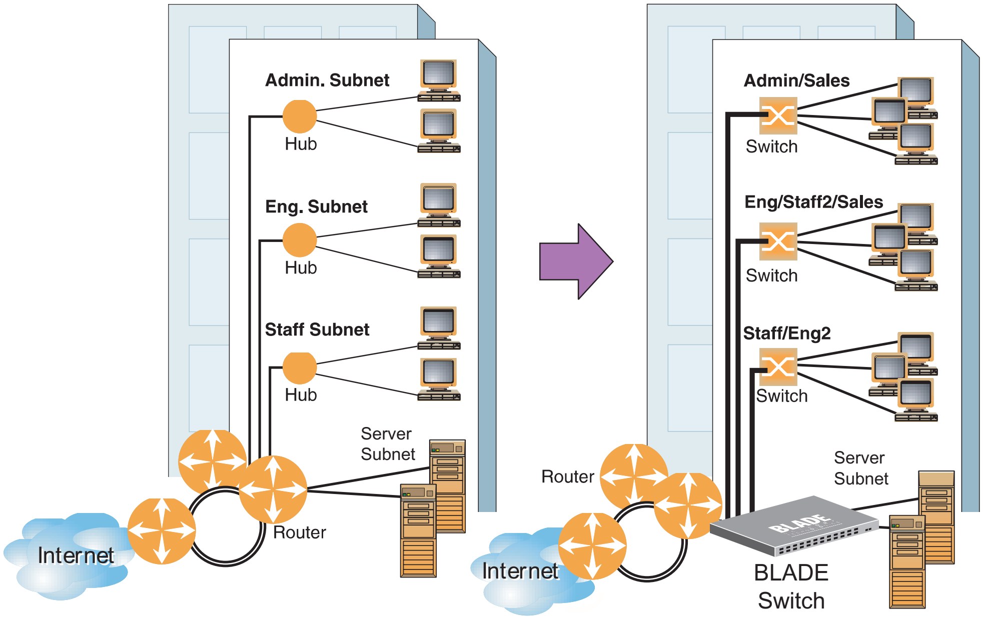
Routing Between IP Subnets

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 G8000 is intelligent and fast enough to perform routing functions on a 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.

For example, consider the following topology migration:

Figure 19 The Router Legacy Network



In this example, 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.

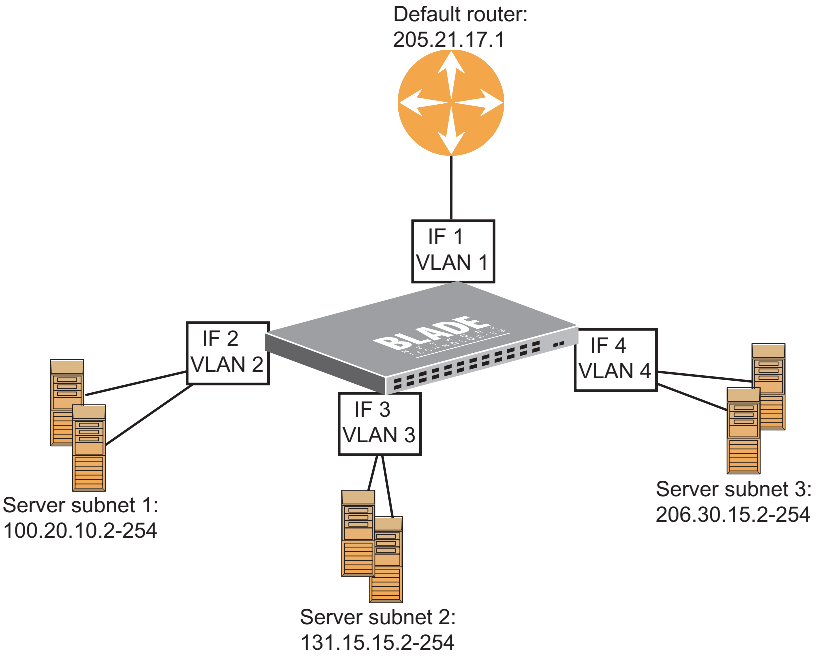
Even if every end-station could be moved to better logical subnets (a daunting task), competition for access to common server pools on different subnets still burdens the routers.

This problem is solved by using switches 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.

Example of Subnet Routing

Consider the role of the G8000 in the following configuration example:

Figure 20 Switch-Based Routing Topology



The switch connects the Gigabit Ethernet and Fast Ethernet trunks 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 switch, 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.

Using VLANs to Segregate Broadcast Domains

If you want to control the broadcasts on your network, use VLANs to create distinct broadcast domains. Create one VLAN for each server subnet, and one for the router.

Configuration Example

This section describes the steps used to configure the example topology shown in Figure 20 on page 211.

1. Assign an IP address (or document the existing one) for each router and each server.

   The following IP addresses are used:

Table 16 Subnet Routing Example: IP Address Assignments

Subnet	Devices	IP Addresses
1	Default router	205.21.17.1
2	Web servers	100.20.10.2-254
3	Database servers	131.15.15.2-254
4	Terminal 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 17 Subnet Routing Example: IP Interface Assignments

Interface	Devices	IP Interface Address
IF 1	Default router	205.21.17.3
IF 2	Web servers	100.20.10.1
IF 3	Database servers	131.15.15.1
IF 4	Terminal Servers	206.30.15.1

3. Determine which switch ports and IP interfaces belong to which VLANs.

   The following table adds port and VLAN information:

Table 18 Subnet Routing Example: Optional VLAN Ports

Devices	IP Interface	Switch Ports	VLAN #
Default router	1	22	1
Web servers	2	1 and 2	2
Database servers	3	3 and 4	3
Terminal Servers	4	5 and 6	4

Note – To perform this configuration, you must be connected to the switch Command Line Interface (CLI) as the administrator.

---

4. Add the switch ports to their respective VLANs.

   The VLANs shown in Table 18 are configured as follows:

```
RS G8000(config)# vlan 1
RS G8000(config-vlan)# member 22          (Add ports to VLAN 1)
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# exit
RS G8000(config)# vlan 2
RS G8000(config-vlan)# member 1,2         (Add ports to VLAN 2)
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# exit
RS G8000(config)# vlan 3
RS G8000(config-vlan)# member 3,4         (Add ports to VLAN 3)
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# exit
RS G8000(config)# vlan 4
RS G8000(config-vlan)# member 5,6         (Add ports to VLAN 4)
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# exit
```

Each time you add a port to a VLAN, you may get the following prompt:

```
Port 4 is an untagged port and its PVID is changed from 1 to 3.
```

5. Assign a VLAN to each IP interface.

   Now that the ports are separated into VLANs, the VLANs are assigned to the appropriate IP interface for each subnet. From Table 18 on page 213, the settings are made as follows:

```
RS G8000(config)# interface ip 1                    (Select IP interface 1)
RS G8000(config-ip-if)# ip address 205.21.17.3
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# vlan 1                      (Add VLAN 1)
RS G8000(config-ip-if)# enable
RS G8000(config-vlan)# exit
RS G8000(config)# interface ip 2                    (Select IP interface 2)
RS G8000(config-ip-if)# ip address 100.20.10.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# vlan 2                      (Add VLAN 2)
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 3                    (Select IP interface 3)
RS G8000(config-ip-if)# ip address 131.15.15.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# vlan 3                      (Add VLAN 3)
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 4                    (Select IP interface 4)
RS G8000(config-ip-if)# ip address 206.30.15.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# vlan 4                      (Add VLAN 4)
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
```

6. Configure the default gateway to the routers’ addresses.

   The default gateway allows the switch to send outbound traffic to the router:

```
RS G8000(config)# ip gateway 1 address 205.21.17.1
RS G8000(config)# ip gateway 1 enable
```

7. Enable IP routing.

```
RS G8000(config)# ip routing
```

8. Verify the configuration.

```
RS G8000(config)# show vlan
RS G8000(config)# show interface information
RS G8000(config)# show interface ip
```

   Examine the resulting information. If any settings are incorrect, make the appropriate changes.

ECMP Static Routes

Equal-Cost Multi-Path (ECMP) is a forwarding mechanism that routes packets along multiple paths of equal cost. ECMP provides equally-distributed link load sharing across the paths. The hashing algorithm used is based on the source IP address (SIP). ECMP routes allow the switch to choose between several next hops toward a given destination. The switch performs periodic health checks (ping) on each ECMP gateway. If a gateway fails, it is removed from the routing table, and an SNMP trap is sent.

OSPF Integration

When a dynamic route is added through Open Shortest Path First (OSPF), the switch checks the route's gateway against the ECMP static routes. If the gateway matches one of the single or ECMP static route destinations, then the OSPF route is added to the list of ECMP static routes. Traffic is load-balanced across all of the available gateways. When the OSPF dynamic route times out, it is deleted from the list of ECMP static routes.

ECMP Route Hashing

You can configure the parameters used to perform ECMP route hashing, as follows:

- `sip`: Source IP address (default)
- `dipsip`: Source IP address and destination IP address

The ECMP hash setting applies to all ECMP routes.

Configuring ECMP Static Routes

To configure ECMP static routes, add the same route multiple times, each with the same destination IP address, but with a different gateway IP address. These routes become ECMP routes.

1. Add a static route (IP address, subnet mask, gateway, and interface number).

```
RS G8000(config)# ip route 10.10.1.1 255.255.255.255 100.10.1.1 1
```

2. Add another static route with the same IP address and mask, but a different gateway address.

```
RS G8000(config)# ip route 10.10.1.1 255.255.255.255 200.20.2.2 1
```

3. Select an ECMP hashing method (optional).

```
RS G8000(config)# ip route ecmphash [sip|dipsip]
```

You may add up to five (5) gateways for each static route.

Use the following command to check the status of ECMP static routes:

```
RS G8000(config)# show ip route static

Current ecmp static routes:
Destination     Mask            Gateway         If    GW Status
--------------- --------------- --------------- ---- -----------
10.10.1.1       255.255.255.255 100.10.1.1        1    up
                                200.20.2.2        1    down

10.20.2.2       255.255.255.255 10.233.3.3        1    up
10.20.2.2       255.255.255.255 10.234.4.4        1    up
10.20.2.2       255.255.255.255 10.235.5.5        1    up

ECMP health-check ping interval: 1
ECMP health-check retries number: 3
ECMP Hash Mechanism: sip
```

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.

The switch accepts gateway configuration parameters if they have not been configured manually. The switch ignores DHCP gateway parameters if the gateway is configured.

DHCP is an extension of another network IP management protocol, Bootstrap Protocol (BOOTP), with an additional capability of being able to 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).

To enable DHCP on a switch interface, use the following command:

```
RS G8000(config)# system dhcp
```

DHCP Relay Agent

DHCP is described in RFC 2131, and the DHCP relay agent supported on the G8000 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 G8000 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.

To enable the G8000 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.

In G8000 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:

```
RS G8000(config)# ip bootp-relay server 1 <IP address>
RS G8000(config)# ip bootp-relay server 2 <IP address>
RS G8000(config)# ip bootp-relay enable
RS G8000(config)# show ip bootp-relay
```

Additionally, DHCP Relay functionality can be assigned on a per interface basis. Use the following commands to enable the Relay functionality:

```
RS G8000(config)# interface ip <Interface number>
RS G8000(config-ip-if)# relay
```

CHAPTER 15

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 2460
- RFC 2461
- RFC 2462
- RFC 2474
- RFC 2526
- RFC 2711
- RFC 3289
- RFC 3411, 3412, 3413, 3414
- RFC 3484
- RFC 3879
- RFC 4007
- RFC 4213
- RFC 4291
- RFC 4293, 4293
- RFC 4443
- RFC 4861
- RFC 4862
- RFC 5095

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)
- Multicast Listener Discovery (MLD)

Most other BLADEOS 6.5 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
- SNMP trap host destination IP address
- Bootstrap Protocol (BOOTP) and DHCP
- RADIUS, TACACS+ and LDAP
- QoS metering and re-marking ACLs for out-profile traffic
- Stacking
- VMware Virtual Center (vCenter) for VMready
- Routing Information Protocol (RIP)
- Internet Group Management Protocol (IGMP)
- Border Gateway Protocol (BGP)
- Protocol Independent Multicast (PIM)
- 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 BLADEOS 6.5 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. Autoconfigured 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

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

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 Autoconfiguration

IPv6 supports the following types of address autoconfiguration:

- **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.

BLADEOS 6.5 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.

You can manually configure two IPv6 addresses for each interface, as follows:

- Initial IPv6 address is a global unicast or anycast address.

  ```
  RS G8000(config)# interface ip <interface number>
  RS G8000(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.

  ```
  RS G8000(config-ip-if)# ipv6 secaddr6 <IPv6 address>
  RS G8000(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.

Use the following commands to configure the IPv6 gateway:

```
RS G8000(config)# ip gateway6 1 address <IPv6 address>
RS G8000(config)# ip gateway6 1 enable
```

IPv6 gateway 1 is reserved for IPv6 data interfaces. IPv6 gateway 4 is the default IPv6 management gateway.

Neighbor Discovery

Neighbor Discovery Overview

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 general advertisements, flags, and interval settings, as well as for defining prefix profiles for router advertisements, is performed on a per-interface basis using the following command path:

```
RS G8000(config)# interface ip <interface number>
RS G8000(config-ip-if)# [no] ipv6 nd ?
RS G8000(config-ip-if)# exit
```

To add or remove entries in the static neighbor cache, use the following command path:

```
RS G8000(config)# [no] ip neighbors ?
```

To manage IPv6 prefix policies, use the following command path:

```
RS G8000(config)# [no] ip prefix-policy ?
```

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 autoconfigured. 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:

```
RS G8000(config)# interface ip <interface number>
RS G8000(config-ip-if)# ip6host
RS G8000(config-ip-if)# exit
```

The G8000 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. 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. 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:

  ```
  RS G8000(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.

Configuration Guidelines

When you configure an interface for IPv6, consider the following guidelines:

- IPv6 only supports static routes.
- Support for subnet router anycast addresses is not available.
- 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

This section provides steps to configure IPv6 on the switch.

IPv6 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.

```
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip6host
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
```

2. Configure the IPv6 default gateway.

```
RS G8000(config)# ip gateway6 1 address
    2001:BA98:7654:BA98:FEDC:1234:ABCD:5412
RS G8000(config)# ip gateway6 1 enable
```

3. Verify the interface address.

```
RS G8000(config)# show interface ip 2
```

IPv6 Example 2

Use the following example to manually configure IPv6 on an interface.

1. Assign an IPv6 address and prefix length to the interface.

```
RS G8000(config)# interface ip 3
RS G8000(config-ip-if)# ipv6 address
    2001:BA98:7654:BA98:FEDC:1234:ABCD:5214
RS G8000(config-ip-if)# ipv6 prefixlen 64
RS G8000(config-ip-if)# ipv6 seccaddr6 2003::1 32
RS G8000(config-ip-if)# vlan 2
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
```

The secondary IPv6 address is compressed, and the prefix length is 32.

2. Configure the IPv6 default gateway.

```
RS G8000(config)# ip gateway6 1 address
    2001:BA98:7654:BA98:FEDC:1234:ABCD:5412
RS G8000(config)# ip gateway6 1 enable
```

3. Configure Neighbor Discovery advertisements for the interface (optional)

```
RS G8000(config)# interface ip 3
RS G8000(config-ip-if)# no ipv6 nd suppress-ra
```

4. Verify the configuration.

```
RS G8000(config-ip-if)# show layer3
```

CHAPTER 16

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). BLADEOS software supports RIP version 1 (RIPv1) and RIP version 2 (RIPv2) for exchanging TCP/IPv4 route information with other routers.

Note – BLADEOS 6.5 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 metric 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 implementing a limit on 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 respective 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 section, Routing Information Protocol Configuration in the *BLADEOS Command Reference*.

RIPv1

RIP version 1 use 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. BLADEOS supports using clear password for RIPv2.

RIPv2 in RIPv1 Compatibility Mode

BLADEOS 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

BLADEOS provides the following features to support RIPv1 and RIPv2:

Poison

Simple split horizon in RIP scheme omits routes learned from one neighbor in updates sent to that neighbor. That is the most common configuration used in RIP, that is setting this Poison to DISABLE. Split horizon with poisoned reverse 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 broadcasts. Multicast RIPv2 announcements 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

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 plaintext 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 messages 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; 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.

```
>> # vlan 2
>> (config-vlan)# enable
>> (config-vlan)# member 2
Port 2 is an UNTAGGED port and its current PVID is 1.
Confirm changing PVID from 1 to 2 [y/n]: y
>> (config-vlan)# exit
>> # vlan 3
>> (config-vlan)# enable
>> (config-vlan)# member 3
Port 3 is an UNTAGGED port and its current PVID is 1.
Confirm changing PVID from 1 to 3 [y/n]: y
>> (config-vlan)# exit
```

2. Add IP interfaces with IPv4 addresses to VLANs.

```
>> # interface ip 2
>> (config-ip-if)# enable
>> (config-ip-if)# address 102.1.1.1
>> (config-ip-if)# vlan 2
>> (config-ip-if)# exit
>> # interface ip 3
>> (config-ip-if)# enable
>> (config-ip-if)# address 103.1.1.1
>> (config-ip-if)# vlan 3
```

3. Turn on RIP globally and enable RIP for each interface.

```
>> # router rip
>> (config-router-rip)# enable
>> (config-router-rip)# exit
>> # interface ip 2
>> (config-ip-if)# ip rip enable
>> (config-ip-if)# exit
>> # interface ip 3
>> (config-ip-if)# ip rip enable
>> (config-ip-if)# exit
```

Use the following command to check the current valid routes in the routing table of the switch:

```
>> # show ip route
```

For those RIP routes learned within the garbage collection period, that are routes phasing out of the routing table with metric 16, use the following command:

```
>> # show ip rip
```

Locally configured static routes do not appear in the RIP Routes table.

CHAPTER 17

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 G8000 can perform IGMP Snooping, and connect to static multicast routers (Mrouters).

Note – BLADEOS 6.5 does not support IPv6 for IGMP.

The following topics are discussed in this chapter:

- "IGMP Snooping" on page 240
- "IGMP Relay" on page 245
- "IGMP Filtering" on page 247

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 a *Leave Report* to the switch, which sends a proxy Leave Report to the Mrouter. The multicast path is terminated immediately.

The G8000 supports the following:

- IGMP version 1 and 2
- IGMP version 3 in stand-alone (non-stacking) mode only
- 128 Mrouters

Note – Unknown multicast traffic is sent to all ports if the flood option is enabled and no Membership Report was learned for that specific IGMP group. To enable or disable IGMP flood, use the following command:

```
RS G8000(config)# [no] ip igmp flood
```

IGMP Groups

The G8000 supports a maximum of 2048 IGMP entries, on a maximum of 1024 VLANs.

When the switch is in stacking mode, one IGMP entry is allocated for each unique join request, based on the combination of the port, VLAN, and IGMP group address. If multiple ports join the same IGMP group, they require separate IGMP entries, even if using the same VLAN.

In stand-alone (non-stacking) mode, 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.

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 the following conditions apply:

- If the switch does not receive an IGMP Membership Report within the query-response-interval.
- If no multicast routers have been learned on that port.

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. To enable FastLeave, use the following command:

```
RS G8000(config)# ip igmp fastleave <VLAN number>
```

IGMPv3 Snooping

IGMPv3 includes new membership report messages to extend IGMP functionality. The switch provides snooping capability for all types of IGMP version 3 (IGMPv3) Membership Reports.

IGMPv3 is supported in stand-alone (non-stacking) mode only.

IGMPv3 supports Source-Specific Multicast (SSM). SSM identifies session traffic by both source and group addresses.

The IGMPv3 implementation keeps records on the multicast hosts present in the network. If a host is already registered, when it sends an IS_INC/TO_INC/IS_EXC/TO_EXC report, the switch overwrites the existing (port-host-group) registration with the new registration; the registrations of other hosts on the same group, same port are not changed. IS_INCLUDE/TO_INCLUDE reports with no source are not registered.

The switch 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:

```
RS G8000(config)# no ip igmp snoop igmpv3 exclude
```

By default, the switch snoops the first eight sources listed in the IGMPv3 Group Record. Use the following command to change the number of snooping sources:

```
RS G8000(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:

```
RS G8000(config)# no ip igmp snoop igmpv3 v1v2
```

IGMP Snooping Configuration Example

This section provides steps to configure IGMP Snooping on the switch.

1. Configure port and VLAN membership on the switch.
2. Add VLANs to IGMP Snooping.

```
RS G8000(config)# ip igmp snoop vlan 1
```

3. Enable IGMP Snooping.

```
RS G8000(config)# ip igmp snoop enable
```

4. Enable IGMPv3 Snooping (optional).

```
RS G8000(config)# ip igmp snoop igmpv3 enable
```

5. Enable the IGMP feature.

```
RS G8000(config)# ip igmp enable
```

6. View dynamic IGMP information.

```
RS G8000# show ip igmp groups

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       INC    2:26    Yes
    *           236.0.0.1        9      1       V3       EXC      -     Yes

RS G8000# show ip igmp mrouter

VLAN      Port       Version   Expires    Max Query Resp. Time  QRV   QQIC
------    -------    --------  --------   --------------------  ---   ----

 1        4           V2       static     -                      -     -
 2        3           V3       4:09       128                    2     125
```

These commands display information about IGMP Groups and Mrouters learned by the switch.

Static Multicast Router

A static multicast router (Mrouter) can be configured for a particular port on a particular VLAN. A static Mrouter does not have to be learned through IGMP Snooping. Any data port can accept a static Mrouter.

When you configure a static Mrouter on a VLAN, it replaces any dynamic Mrouters learned through IGMP Snooping.

Configure a Static Multicast Router

1. For each MRouter, configure a port (1-48), VLAN (1-4094) and version (1-3).

```
RS G8000(config)# ip igmp mrouter 5 1 2
```

   The IGMP version is set for each VLAN, and cannot be configured separately for each Mrouter.

2. Verify the configuration.

```
RS G8000# show ip igmp mrouter
```

IGMP Relay

The G8000 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 G8000 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 G8000, 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 G8000 uses health checks to select the primary Mrouter.

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 are mutually exclusive—if you enable IGMP Relay, you must turn off IGMP Snooping.
- Add the upstream Mrouter VLAN to the IGMP Relay list, using the following command:

```
RS G8000(config)# ip igmp relay vlan <VLAN number>
```

- If IGMP hosts reside on different VLANs, you must:
    - Disable IGMP flooding.

    ```
    RS G8000(config)# no ip igmp flood
    ```

    - Enable CPU forwarding to ensure that multicast data is forwarded across the VLANs.

    ```
    RS G8000(config)# ip igmp cpu
    ```

Configure IGMP Relay

Use the following procedure to configure IGMP Relay.

1. Configure IP interfaces with IPv4 addresses, and assign VLANs.

```
>> # interface ip 2
>> (config-ip-if)# ip address 10.10.1.1
>> (config-ip-if)# ip netmask 255.255.255.0
>> (config-ip-if)# vlan 2
>> (config-ip-if)# enable
>> (config-ip-if)# exit
>> # interface ip 3
>> (config-ip-if)# ip address 10.10.2.1
>> (config-ip-if)# ip netmask 255.255.255.0
>> (config-ip-if)# vlan 3
>> (config-ip-if)# enable
>> (config-ip-if)# exit
```

2. Turn IGMP on.

```
>> # ip igmp enable
```

3. Enable IGMP Relay and add VLANs to the downstream network.

```
>> # ip igmp relay enable
>> # ip igmp relay vlan 2
>> # ip igmp relay vlan 3
```

4. Configure the upstream Mrouters with IPv4 addresses.

```
>> # ip igmp relay mrouter 1 address 100.0.1.2
>> # ip igmp relay mrouter 1 enable
>> # ip igmp relay mrouter 2 address 100.0.2.4
>> # ip igmp relay mrouter 2 enable
```

IGMP Filtering

With IGMP Filtering, you can allow or deny a port to send and receive multicast traffic to certain multicast groups. Unauthorized users are restricted from streaming multicast traffic across the network.

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.

Configure IGMP Filtering

1. Enable IGMP Filtering on the switch.

```
>> # ip igmp filtering
```

2. Define an IGMP filter with IPv4 information.

```
>> # ip igmp profile 1 range 224.0.0.0 226.0.0.0
>> # ip igmp profile 1 action deny
>> # ip igmp profile 1 enable
```

3. Assign the IGMP filter to a port.

```
>> # interface port 3
>> (config-if)# ip igmp profile 1
>> (config-if)# ip igmp filtering
```

CHAPTER 18

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.

RackSwitch G8000s can advertise their IP interfaces and IPv4 addresses using BGP and take BGP feeds from as many as 16 BGP router peers. This allows more resilience and flexibility in balancing traffic from the Internet.

Note – BLADEOS 6.5 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.

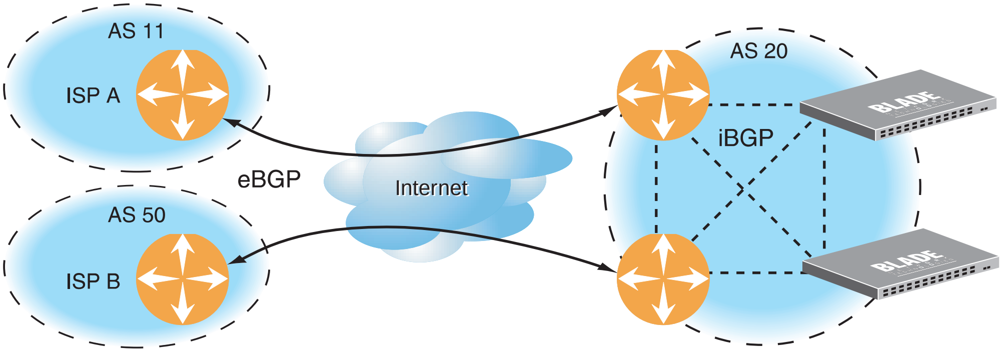
Static routes should have a higher degree of precedence than dynamic routing protocols. If the destination route is not in the route cache, then the packets are forwarded to the default gateway which may be incorrect if a dynamic routing protocol is enabled.

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 21), 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 RackSwitch G8000.

Figure 21 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?

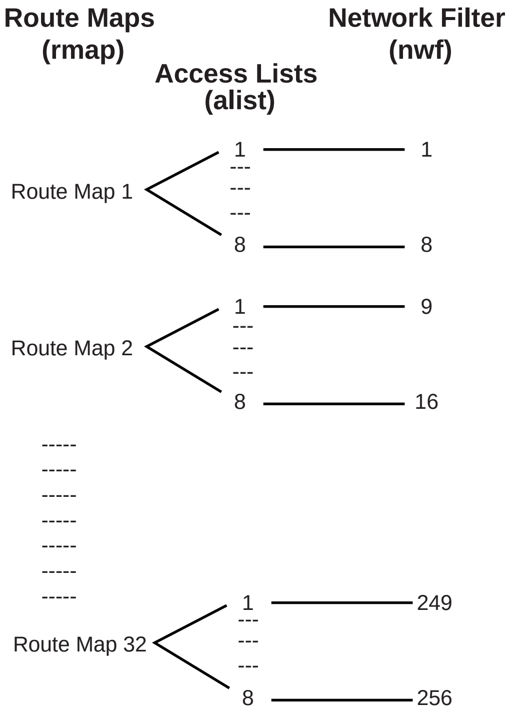
A route map is used to control and modify routing information. Route maps define conditions for redistributing routes from one routing protocol to another or controlling routing information when injecting it in and out of BGP. Route maps are used by OSPF only for redistributing routes. For example, a 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 command is used to enter the Route Map mode for defining a route map:

```
RS G8000(config)# route-map <map number>       (Select a route map)
RS G8000(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 258.

BLADEOS 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 22 illustrates the relationship between route maps, access lists and network filters.

Figure 22 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 command (Route Map mode):

```
RS G8000(config)# route-map <map number>                (Select a route map)
RS G8000(config-route-map)# precedence <1-255>          (Specify a precedence)
RS G8000(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 Overview

To configure route maps, you need to do the following:

1. Define a network filter.

```
RS G8000(config)# ip match-address 1 <IPv4 address> <IPv4 subnet mask>
RS G8000(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.

```
RS G8000(config)# route-map 1
RS G8000(config-route-map)# access-list 1 match-address 1
RS G8000(config-route-map)# access-list 1 metric <metric value>
RS G8000(config-route-map)# access-list 1 action deny
RS G8000(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 AS filter attributes.

```
RS G8000(config-route-map)# as-path-list 1 as 1
RS G8000(config-route-map)# as-path-list 1 action deny
RS G8000(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 the 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.

```
RS G8000(config-route-map)# as-path-preference <AS number>
RS G8000(config-route-map)# local-preference <local preference number>
RS G8000(config-route-map)# metric <metric value>
```

5. Enable the route map.

```
RS G8000(config-route-map)# enable
RS G8000(config-route-map)# exit
```

6. Turn BGP on.

```
RS G8000(config)# router bgp
RS G8000(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,

```
RS G8000(config-router-bgp)# neighbor 1 route-map in <1-32>
```

   or to the outgoing route map list.

```
RS G8000(config-router-bgp)# neighbor 1 route-map out <1-32>
```

8. Exit Router BGP mode.

```
RS G8000(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:

```
>> # router bgp
>> (config-router-bgp)# aggregate-address <1-16> <IPv4 address> <mask>
>> (config-router-bgp)# aggregate-address <1-16> enable
```

An example of creating a BGP aggregate route is shown in "Default Redistribution and Route Aggregation Example" on page 260.

Redistributing Routes

In addition to running multiple routing protocols simultaneously, BLADEOS 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 251. 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 260.

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:

- The following commands use the BGP default local preference method, affecting the outbound direction only.

```
>> # router bgp
>> (config_router_bgp)# local-preference
>> (config_router_bgp)# exit
```

- The following commands use the route map local preference method, which affects both inbound and outbound directions.

```
>> # route-map 1
>> (config_route_map)# local-preference
>> (config_router_map)# exit
```

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 G8000 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 are from eBGP, the route with the lower router ID is selected.

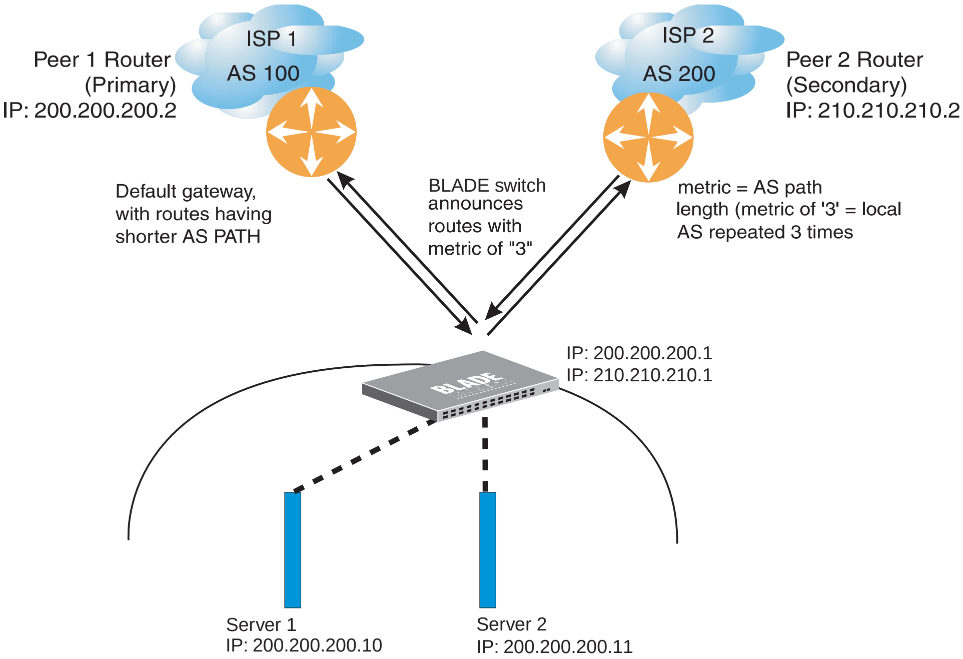
When the path is selected, BGP puts the selected path in its routing table and propagates the path to its neighbors.

BGP Failover Configuration

Use the following example to create redundant default gateways for a G8000 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 23, 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 G8000.

Figure 23 BGP Failover Configuration Example



On the G8000, 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.

```
>> # vlan 1
>> (config-vlan)# member <port number>
```

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.

```
>> # interface ip 1
>> (config-ip-if)# ip address 200.200.200.1
>> (config-ip-if)# ip netmask 255.255.255.0
>> (config-ip-if)# enable
>> (config-ip-if)# exit
>> # interface ip 2
>> (config-ip-if)# ip address 210.210.210.1
>> (config-ip-if)# ip netmask 255.255.255.0
>> (config-ip-if)# enable
>> (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).

```
>> # ip routing
```

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 with IPv4 addresses.

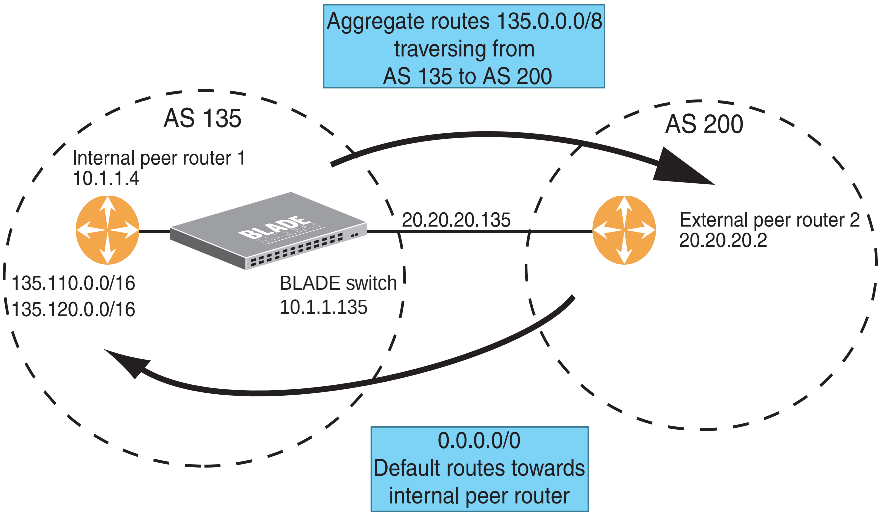
```
>> # router bgp
>> (config-router-bgp)# neighbor 1 remote-address 200.200.200.2
>> (config-router-bgp)# neighbor 1 remote-as 100
>> (config-router-bgp)# neighbor 2 remote-address 210.210.210.2
>> (config-router-bgp)# neighbor 2 remote-as 200
```

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 24, you have two peer routers: an internal and an external peer router. Configure the G8000 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 24 Route Aggregation and Default Route Redistribution



1. Configure the IP interface.
2. Configure the AS number (AS 135) and router ID (10.1.1.135).

```
>> # router bgp
>> (config-router-bgp)# as 135
>> (config-router-bgp)# exit
>> # ip router-id 10.1.1.135
```

3. Configure internal peer router 1 and external peer router 2 with IPv4 addresses.

```
>> # router bgp
>> (config-router-bgp)# neighbor 1 remote-address 10.1.1.4
>> (config-router-bgp)# neighbor 1 remote-as 135
>> (config-router-bgp)# neighbor 2 remote-address 20.20.20.2
>> (config-router-bgp)# neighbor 2 remote-as 200
```

4. Configure redistribution for Peer 1.

```
>> (config-router-bgp)# neighbor 1 redistribute default-action
    redistribute
>> (config-router-bgp)# neighbor 1 redistribute fixed
```

5. Configure aggregation policy control.

   Configure the IPv4 routes that you want aggregated.

```
>> (config-router-bgp)# aggregate-address 1 135.0.0.0 255.0.0.0
>> (config-router-bgp)# aggregate-address 1 enable
```

CHAPTER 19

OSPF

BLADEOS supports the Open Shortest Path First (OSPF) routing protocol. The BLADEOS implementation conforms to the OSPF version 2 specifications detailed in Internet RFC 1583, and OSPF version 3 specifications in RFC 2740. The following sections discuss OSPF support for the RackSwitch G8000:

- "OSPFv2 Overview" on page 263. 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 BLADEOS" on page 268. This section describes how OSPFv2 is implemented in BLADEOS, such as configuration parameters, electing the designated router, summarizing routes, defining route maps and so forth.
- "OSPFv2 Configuration Examples" on page 279. 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 BLADEOS" on page 289. 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.

The following sections describe key OSPF concepts.

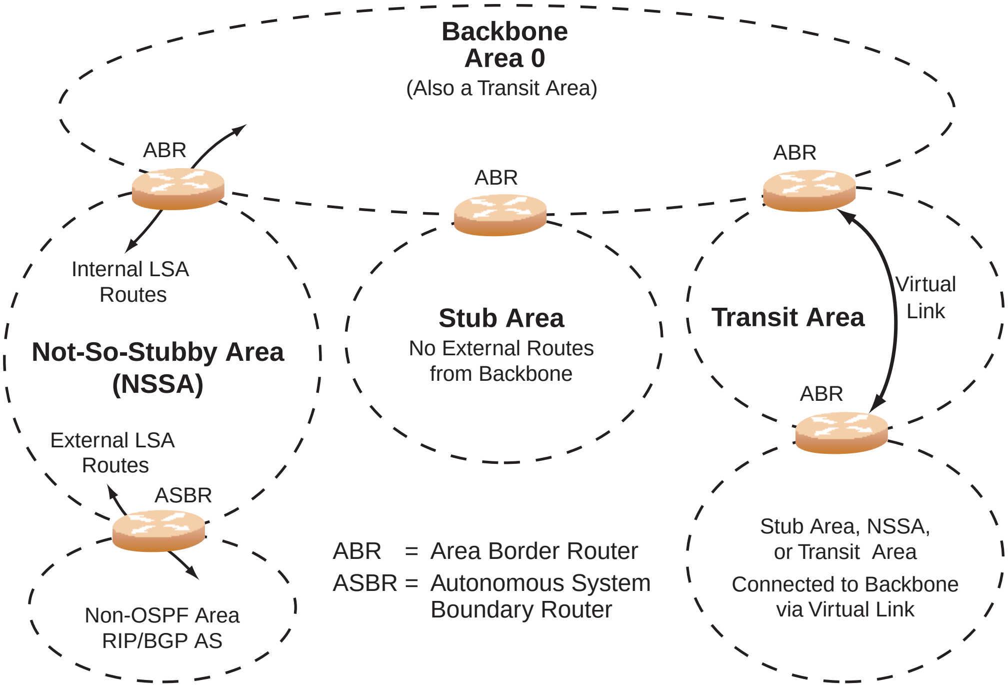
Types of OSPF Areas

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 25, 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 are not distributed into other areas.
- Transit Area—an area that allows area summary information to be exchanged between routing devices. The backbone (area 0), any area that contains a virtual link to connect two areas, and any area that is not a stub area or an NSSA are considered transit areas.

Figure 25 OSPF Area Types

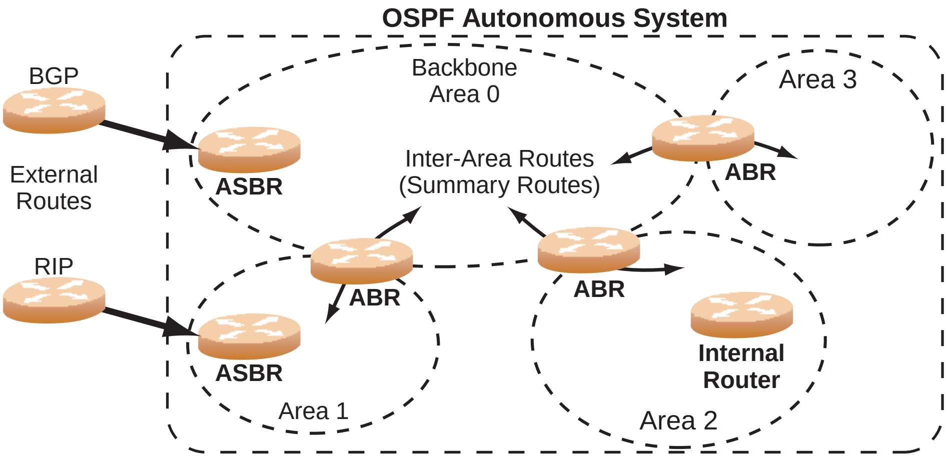


Types of OSPF Routing Devices

As shown in Figure 26, 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 26 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 BLADEOS

BLADEOS supports a single instance of OSPF and up to 4K routes on the network. The following sections describe OSPF implementation in BLADEOS:

- "Configurable Parameters" on page 268
- "Defining Areas" on page 269
- "Interface Cost" on page 271
- "Electing the Designated Router and Backup" on page 271
- "Summarizing Routes" on page 271
- "Default Routes" on page 272
- "Virtual Links" on page 273
- "Router ID" on page 274
- "Authentication" on page 275

Configurable Parameters

In BLADEOS, 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 25.

The ISCLI supports the following parameters: interface output cost, interface priority, dead and hello intervals, retransmission interval, and interface transmit delay.

In addition to the above parameters, you can also 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 G8000 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 273).

Up to six OSPF areas can be connected to the G8000 with BLADEOS 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:

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# area <area index> area-id <n.n.n.n>
RS G8000(config-router-ospf)# area <area index> enable
RS G8000(config-router-ospf)# exit
```

Note – The `area` 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 `area` portion of the command as explained in the following sections.

Assigning the Area Index

The `area` <*area index*> option is actually just an arbitrary index (0–5) used only by the G8000. This index number 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 in order 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:

- Placing the area number in the last octet (0.0.0.*n*)

  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.`" On the G8000, using the last octet in the area ID, "`area 1`" is equivalent to "`area-id 0.0.0.1`".

- Multi-octet (*IP address*)

  Some OSPF vendors express the area ID number in multi-octet format. For example, "`area 2.2.2.2`" represents OSPF area 2 and can be specified directly on the G8000 as "`area-id 2.2.2.2`".

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 format for the command is as follows:

```
RS G8000(config)# interface ip <interface number>
RS G8000(config-ip-if)# ip ospf area <area index>
RS G8000(config-ip-if)# exit
```

For example, the following commands could be used to configure IP interface 14 for a presence on the 10.10.10.1/24 network, to define OSPF area 1, and to attach the area to the network:

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# area 1 area-id 0.0.0.1
RS G8000(config-router-ospf)# enable
RS G8000(config-router-ospf)# exit
RS G8000(config)# interface ip 14
RS G8000(config-ip-if)# ip address 10.10.10.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# ip ospf area 1
RS G8000(config-ip-if)# ip ospf enable
```

Note – OSPFv2 supports IPv4 only. IPv6 is supported in OSPFv3 (see "OSPFv3 Implementation in BLADEOS" on page 289).

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 (Interface IP mode):

```
RS G8000(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 G8000. The command is as follows:

```
RS G8000(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:

```
RS G8000(config-ip-if)# ip ospf passive-interface
RS G8000(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:

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# area-range <range number> address
        <IP address> <mask>
```

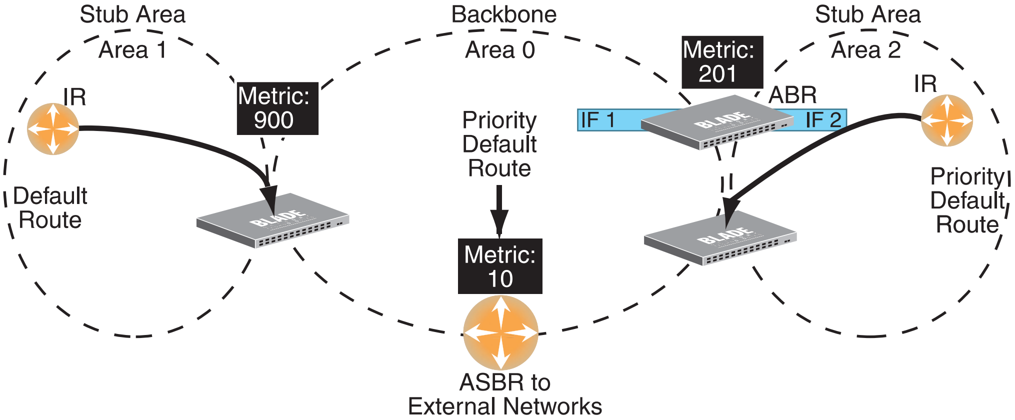
where *<range number>* is a number 1 to 16, *<IP address>* is the base IP address for the range, and *<mask>* is the IP address mask for the range. For a detailed configuration example, see "Example 3: Summarizing Routes" on page 286.

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 G8000 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 27), 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 27 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 (Router OSPF mode):

```
RS G8000(config-router-ospf)# default-information <metric value>
    <metric type (1 or 2)>
```

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:

```
RS G8000(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 25 on page 264).

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:

```
RS G8000(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 G8000 as one endpoint of a virtual link, use the following command:

```
RS G8000(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, the routing device at the target endpoint. Another router ID is needed when configuring a virtual link in the other direction. To provide the G8000 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 282.

Router ID

Routing devices in OSPF areas are identified by a router ID. The router ID is expressed in IP address format. The IP address of the router ID is not required to be included in any IP interface range or in any OSPF area, and may even use the G8000 loopback interface.

The router ID can be configured in one of the following two ways:

- Dynamically—OSPF protocol configures the lowest IP interface IP address as the router ID (loopback interface has priority over the IP interface). This is the default.
- Statically—Use the following command to manually configure the router ID:

  ```
  RS G8000(config-router-ospf)# ip router-id <IPv4 address>
  ```

  If there is a loopback interface, its IP address is always preferred as the router ID, instead of an IP interface address. The **`ip router-id`** command is the preferred method to set the router ID and it is always used in preference to the other methods.
- To modify the router ID from static to dynamic, set the router ID to 0.0.0.0, save the configuration, and reboot the G8000.
- To view the router ID, use the following command:

  ```
  RS G8000(config-router-ospf)# show ip ospf
  ```

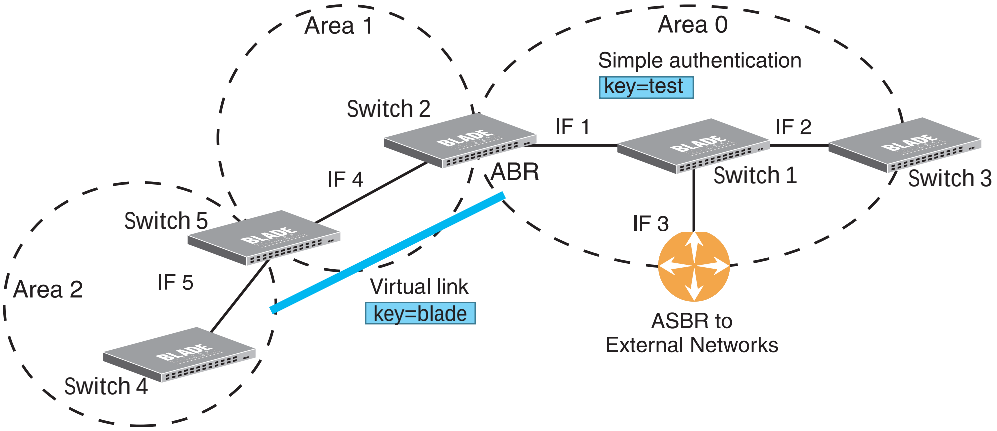
Authentication

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. BLADEOS supports simple password (type 1 plain text passwords) and MD5 cryptographic authentication. This type of authentication allows a password to be configured per area.

Figure 28 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 28 OSPF Authentication



Configuring Plain Text OSPF Passwords

To configure simple plain text OSPF passwords on the switches shown in Figure 28 use the following commands:

1. Enable OSPF authentication for Area 0 on switches 1, 2, and 3.

```
RS G8000(config-router-ospf)# area 0 authentication-type password
RS G8000(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.

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip ospf key test
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip ospf key test
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 3
RS G8000(config-ip-if)# ip ospf key test
RS G8000(config-ip-if)# exit
```

3. Enable OSPF authentication for Area 2 on switch 4.

```
RS G8000(config)# router ospf
RS G8000(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.

```
RS G8000(config-router-ospf)# area-virtual-link 1 key blade
```

Configuring MD5 Authentication

Use the following commands to configure MD5 authentication on the switches shown in Figure 28:

1. Enable OSPF MD5 authentication for Area 0 on switches 1, 2, and 3.

```
RS G8000(config-router-ospf)# area 0 authentication-type md5
```

2. Configure MD5 key ID for Area 0 on switches 1, 2, and 3.

```
RS G8000(config-router-ospf)# message-digest-key 1 md5-key test
RS G8000(config-router-ospf)# exit
```

3. Assign MD5 key ID to OSPF interfaces on switches 1, 2, and 3.

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip ospf message-digest-key 1
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip ospf message-digest-key 1
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 3
RS G8000(config-ip-if)# ip ospf message-digest-key 1
RS G8000(config-ip-if)# exit
```

4. Enable OSPF MD5 authentication for Area 2 on switch 4.

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# area 1 authentication-type md5
```

5. Configure MD5 key for the virtual link between Area 2 and Area 0 on switches 2 and 4.

```
RS G8000(config-router-ospf)# message-digest-key 2 md5-key test
```

6. Assign MD5 key ID to OSPF virtual link on switches 2 and 4.

```
RS G8000(config-router-ospf)# area-virtual-link 1 message-digest-key 2
RS G8000(config-router-ospf)# exit
```

Host Routes for Load Balancing

BLADEOS 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 274) 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.

Note – Loopback interfaces are not advertised via the OPSF route redistribution of fixed routes.

OSPF Features Not Supported in This Release

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 G8000 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.

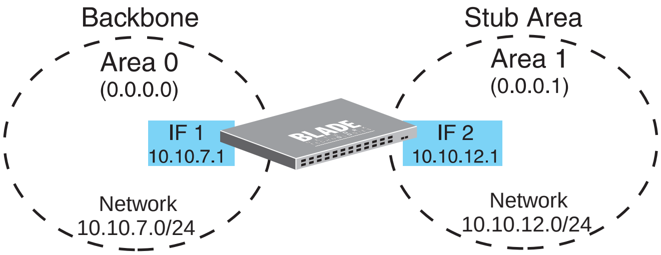
   IP interfaces are used for attaching networks to the various areas.

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 29 A Simple OSPF Domain



Follow this procedure to configure OSPF support as shown in Figure 29:

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

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip address 10.10.7.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip address 10.10.12.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
```

Note – OSPFv2 supports IPv4 only. IPv6 is supported in OSPFv3 (see "OSPFv3 Implementation in BLADEOS" on page 289).

2. Enable OSPF.

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# enable
```

3. Define the backbone.

   The backbone is always configured as a transit area using `areaid 0.0.0.0`.

```
RS G8000(config-router-ospf)# area 0 area-id 0.0.0.0
RS G8000(config-router-ospf)# area 0 type transit
RS G8000(config-router-ospf)# area 0 enable
```

4. Define the stub area.

```
RS G8000(config-router-ospf)# area 1 area-id 0.0.0.1
RS G8000(config-router-ospf)# area 1 type stub
RS G8000(config-router-ospf)# area 1 enable
RS G8000(config-router-ospf)# exit
```

5. Attach the network interface to the backbone.

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip ospf area 0
RS G8000(config-ip-if)# ip ospf enable
RS G8000(config-ip-if)# exit
```

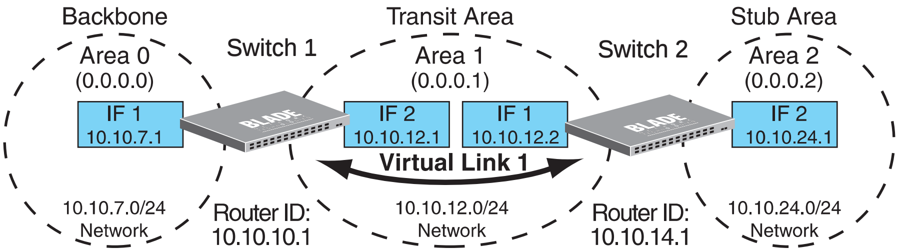
6. Attach the network interface to the stub area.

```
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip ospf area 1
RS G8000(config-ip-if)# ip ospf enable
RS G8000(config-ip-if)# exit
```

Example 2: Virtual Links

In the example shown in Figure 30, 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 30 Configuring a Virtual Link



Note – OSPFv2 supports IPv4 only. IPv6 is supported in OSPFv3 (see "OSPFv3 Implementation in BLADEOS" on page 289).

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

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip address 10.10.7.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip address 10.10.12.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(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.

```
RS G8000(config)# ip router-id 10.10.10.1
```

3. Enable OSPF.

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# enable
```

4. Define the backbone.

```
RS G8000(config-router-ospf)# area 0 area-id 0.0.0.0
RS G8000(config-router-ospf)# area 0 type transit
RS G8000(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.

```
RS G8000(config-router-ospf)# area 1 area-id 0.0.0.1
RS G8000(config-router-ospf)# area 1 type transit
RS G8000(config-router-ospf)# area 1 enable
RS G8000(config-router-ospf)# exit
```

6. Attach the network interface to the backbone.

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip ospf area 0
RS G8000(config-ip-if)# ip ospf enable
RS G8000(config-ip-if)# exit
```

7. Attach the network interface to the transit area.

```
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip ospf area 1
RS G8000(config-ip-if)# ip ospf enable
RS G8000(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 284.

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# area-virtual-link 1 area 1
RS G8000(config-router-ospf)# area-virtual-link 1 neighbor-router
     10.10.14.1
RS G8000(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

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip address 10.10.12.2
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip address 10.10.24.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(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 283.

```
RS G8000(config)# ip router-id 10.10.14.1
```

3. Enable OSPF.

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# enable
```

4. Define the backbone.

   This version of BLADEOS requires that a backbone index be configured on the non-backbone end of the virtual link as follows:

```
RS G8000(config-router-ospf)# area 0 area-id 0.0.0.0
RS G8000(config-router-ospf)# area 0 enable
```

5. Define the transit area.

```
RS G8000(config-router-ospf)# area 1 area-id 0.0.0.1
RS G8000(config-router-ospf)# area 1 type transit
RS G8000(config-router-ospf)# area 1 enable
```

6. Define the stub area.

```
RS G8000(config-router-ospf)# area 2 area-id 0.0.0.2
RS G8000(config-router-ospf)# area 1 type stub
RS G8000(config-router-ospf)# area 1 enable
RS G8000(config-router-ospf)# exit
```

7. Attach the network interface to the backbone.

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip ospf area 1
RS G8000(config-ip-if)# ip ospf enable
RS G8000(config-ip-if)# exit
```

8. Attach the network interface to the transit area.

```
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip ospf area 2
RS G8000(config-ip-if)# ip ospf enable
RS G8000(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 282.

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# area-virtual-link 1 area 1
RS G8000(config-router-ospf)# area-virtual-link 1 neighbor-router
     10.10.10.1
RS G8000(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

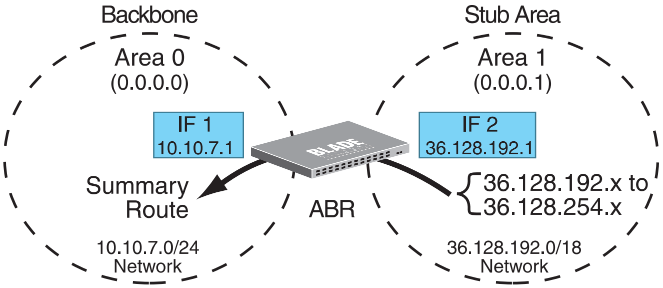
By default, ABRs advertise all the network addresses from one area into another area. Route summarization can be used for consolidating advertised addresses and reducing the perceived complexity of the network.

If 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 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 BLADEOS" on page 289).

Figure 31 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.

Use the following procedure to configure OSPF support as shown in Figure 31:

1. Configure IP interfaces for each network which will be attached to OSPF areas.

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip address 10.10.7.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip address 36.128.192.1
RS G8000(config-ip-if)# ip netmask 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
```

2. Enable OSPF.

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# enable
```

3. Define the backbone.

```
RS G8000(config-router-ospf)# area 0 area-id 0.0.0.0
RS G8000(config-router-ospf)# area 0 type transit
RS G8000(config-router-ospf)# area 0 enable
```

4. Define the stub area.

```
RS G8000(config-router-ospf)# area 1 area-id 0.0.0.1
RS G8000(config-router-ospf)# area 1 type stub
RS G8000(config-router-ospf)# area 1 enable
RS G8000(config-router-ospf)# exit
```

5. Attach the network interface to the backbone.

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip ospf area 0
RS G8000(config-ip-if)# ip ospf enable
RS G8000(config-ip-if)# exit
```

6. Attach the network interface to the stub area.

```
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip ospf area 1
RS G8000(config-ip-if)# ip ospf enable
RS G8000(config-ip-if)# exit
```

7. Configure route summarization by specifying the starting address and mask of the range of addresses to be summarized.

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# area-range 1 address 36.128.192.0
      255.255.192.0
RS G8000(config-router-ospf)# area-range 1 area 0
RS G8000(config-router-ospf)# area-range 1 enable
RS G8000(config-router-ospf)# exit
```

8. Use the hide command to prevent a range of addresses from advertising to the backbone.

```
RS G8000(config)# router ospf
RS G8000(config-router-ospf)# area-range 2 address 36.128.200.0
      255.255.255.0
RS G8000(config-router-ospf)# area-range 2 area 0
RS G8000(config-router-ospf)# area-range 2 hide
RS G8000(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 *BLADEOS Command Reference* for information on the above commands.

OSPFv3 Implementation in BLADEOS

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 G8000. 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 219 for configuring IPv6 interfaces.

OSPFv3 Uses Independent Command Paths

Though OSPFv3 and OSPFv2 are very similar, they are configured independently. They each have their own separate menus in the CLI, and their own command paths in the ISCLI. OSPFv3 base menus and command paths are located as follows:

- In the CLI

```
>> # /cfg/l3/ospf3                                (OSPFv3 config menu)
>> # /info/l3/ospf3                               (OSPFv3 information menu)
>> # /stats/l3/ospf3                              (OSPFv3 statistics menu)
```

- In the ISCLI

```
RS G8000(config)# ipv6 router ospf                (OSPFv3 router config mode)
RS G8000(config-router-ospf3)# ?

RS G8000(config)# interface ip <Interface number> (Configure OSPFv3)
RS G8000(config-ip-if)# ipv6 ospf ?               (OSPFv3 interface config)

RS G8000# 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.
- Link-local flooding scope has been added, along with a Link LSA. This allows flooding information to relevant local neighbors without forwarded it beyond the local router.
- Flexible treatment of unknown LSA types to make integration of OSPFv3 easier.

OSPFv3 Limitations

BLADEOS 6.5 does not currently support the following OSPFv3 features:

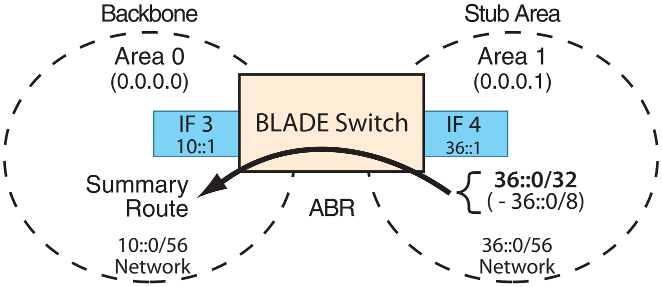
- Multiple instances of OSPFv3 on one IPv6 link.
- Authentication via IPv6 Security (IPsec)

OSPFv3 Configuration Example

The following example depicts the OSPFv3 equivalent configuration of "Example 3: Summarizing Routes" on page 286 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 from the 36::0/32 portion of the 36::0/56 network, except for the routes in the 36::0/8 range.

Figure 32 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 31:

1. Configure IPv6 interfaces for each link which will be attached to OSPFv3 areas.

```
RS G8000(config)# interface ip 3
RS G8000(config-ip-if)# ipv6 address 10:0:0:0:0:0:0:1
RS G8000(config-ip-if)# ipv6 prefixlen 56
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 4
RS G8000(config-ip-if)# ip address 36:0:0:0:0:0:1
RS G8000(config-ip-if)# ipv6 prefixlen 56
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
```

This is equivalent to configuring the IP address and netmask for IPv4 interfaces.

2. Enable OSPFv3.

```
RS G8000(config)# ipv6 router ospf
RS G8000(config-router-ospf3)# enable
```

This is equivalent to the OSPFv2 `enable` option in the `router ospf` command path.

3. Define the backbone.

```
RS G8000(config-router-ospf3)# area 0 area-id 0.0.0.0
RS G8000(config-router-ospf3)# area 0 type transit
RS G8000(config-router-ospf3)# area 0 enable
```

This is identical to OSPFv2 configuration.

4. Define the stub area.

```
RS G8000(config-router-ospf3)# area 1 area-id 0.0.0.1
RS G8000(config-router-ospf3)# area 1 type stub
RS G8000(config-router-ospf3)# area 1 enable
RS G8000(config-router-ospf3)# exit
```

This is identical to OSPFv2 configuration.

5. Attach the network interface to the backbone.

```
RS G8000(config)# interface ip 3
RS G8000(config-ip-if)# ipv6 ospf area 0
RS G8000(config-ip-if)# ipv6 ospf enable
RS G8000(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.

```
RS G8000(config)# interface ip 4
RS G8000(config-ip-if)# ipv6 ospf area 1
RS G8000(config-ip-if)# ipv6 ospf enable
RS G8000(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.

```
RS G8000(config)# ipv6 router ospf
RS G8000(config-router-ospf3)# area-range 1 address 36:0:0:0:0:0:0:0 32
RS G8000(config-router-ospf3)# area-range 1 area 0
RS G8000(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.

```
RS G8000(config-router-ospf)# area-range 2 address 36:0:0:0:0:0:0:0 8
RS G8000(config-router-ospf)# area-range 2 area 0
RS G8000(config-router-ospf)# area-range 2 hide
RS G8000(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.

CHAPTER 20
Protocol Independent Multicast

BLADEOS supports Protocol Independent Multicast (PIM) in Sparse Mode (PIM-SM) and Dense Mode (PIM-DM).

Note – BLADEOS 6.5 does not support IPv6 for PIM.

The following sections discuss PIM support for the RackSwitch G8000:

- "PIM Overview" on page 295
- "Supported PIM Modes and Features" on page 296
- "Basic PIM Settings" on page 297
- "Additional Sparse Mode Settings" on page 300
- "Using PIM with Other Features" on page 302
- "PIM Configuration Examples" on page 303

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 BLADEOS 6.5:

- 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 – In BLADEOS 6.5, 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 commands:

```
RS G8000(config)# [no] ip pim enable
```

Defining a PIM Network Component

The G8000 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:

```
RS G8000(config)# ip pim component <1-2>
RS G8000(config-ip-pim-comp)# mode {sparse|dense}
RS G8000(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 unconfigured 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 300).

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, and they may also belong to different VLANs as long as their member IP addresses do not overlap.

To define an IP interface for use with PIM, first configured the interface with an IPv4 address and VLAN as follows:

```
RS G8000(config)# interface ip <Interface number>
RS G8000(config-ip-if)# ip address <IPv4 address> <IPv4 mask>
RS G8000(config-ip-if)# vlan <VLAN number>
RS G8000(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:

```
RS G8000(config-ip-if)# ip pim enable
RS G8000(config-ip-if)# ip pim component-id <1-2>
RS G8000(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 G8000 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. 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:

```
RS G8000(config)# interface ip <Interface number>
RS G8000(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:

```
RS G8000(config-ip-if)# ip pim neighbor-addr <neighbor IPv4 address> allow
```

To remove a PIM neighbor from the accepted list, use the following command.

```
RS G8000(config-ip-if)# ip pim neighbor-addr <neighbor IPv4 address> deny
RS G8000(config-ip-if)# exit
```

You can view configured PIM neighbor filters globally or for a specific IP interface using the following commands:

```
RS G8000(config)# show ip pim neighbor-filters
RS G8000(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 G8000 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:

```
RS G8000(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:

```
RS G8000(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 301).

Alternately, if no election is desired, the switch can provide a static RP, specified using the following command:

```
RS G8000(config-ip-pim-comp)# rp-static rp-address <group address>
    <group address mask>  <candidate 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:

```
RS G8000(config-ip-pim-comp)# rp-candidate holdtime <1-255>
RS G8000(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):

```
RS G8000(config)# interface ip <Interface number>
RS G8000(config-ip-if)# ip pim dr-priority <value (0-4294967294)>
RS G8000(config-ip-if)# exit
```

Note – A value of 0 (zero) specifies that the G8000 will not act as the DR. This setting requires the G8000 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:

```
RS G8000(config)# interface ip <Interface number>
RS G8000(config-ip-if)# ip pim cbsr-preference <-1 to 255>
RS G8000(config-ip-if)# exit
```

A value of 255 highly prefers the local interface as a BSR. A value of -1 indicates that the local interface should not act as a BSR.

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 239):

- IGMP static joins can be configured with a PIM-SM or PIM-DM multicast group IPv4 address. Using the ISCLI:

  ```
  RS G8000(config)# ip mroute <multicast group IPv4 address> <VLAN> <port>
  ```

  Using the CLI

  ```
  >> # /cfg/l3/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:

```
RS G8000(config)# ip pim enable
```

2. Configure a PIM network component with dynamic RP settings, and set it for PIM Sparse Mode:

```
RS G8000(config)# ip pim component 1
RS G8000(config-ip-pim-comp)# mode sparse
RS G8000(config-ip-pim-comp)# rp-candidate rp-address 225.1.0.0
    255.255.0.0 10.10.1.1
RS G8000(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:

```
RS G8000(config)# interface ip 111
RS G8000(config-ip-if)# ip address 10.10.1.1 255.255.255.255
RS G8000(config-ip-if)# vlan 11
RS G8000(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:

```
RS G8000(config-ip-if)# ip pim enable
RS G8000(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:

```
RS G8000(config-ip-if)# ip pim cbsr-preference 135
RS G8000(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:

```
RS G8000(config)# ip pim static-rp enable
RS G8000(config)# ip pim component 1
RS G8000(config-ip-pim-comp)# rp-static rp-address 225.1.0.0
    255.255.0.0 10.10.1.1
RS G8000(config-ip-pim-comp)# exit
```

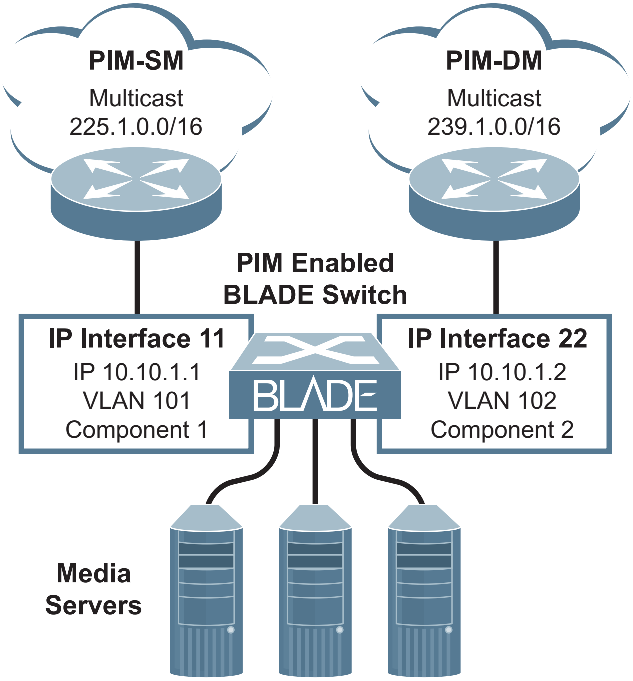
Where 225.1.0.0 255.255.0.0 is the multicast group base address and mask, and 10.10.1.1 is the RP candidate address.

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 33.

Figure 33 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.

```
RS G8000(config)# ip pim enable
```

2. Configure a PIM component and set the PIM mode:

```
RS G8000(config)# ip pim component 2
RS G8000(config-ip-pim-comp)# mode dense
RS G8000(config-ip-pim-comp)# exit
```

3. Define an IP interface for use with PIM:

```
RS G8000(config)# interface ip 102
RS G8000(config-ip-if)# ip address 10.10.1.2 255.255.255.255
RS G8000(config-ip-if)# vlan 22
RS G8000(config-ip-if)# enable
```

4. Enable PIM on the IP interface and assign the PIM component:

```
RS G8000(config-ip-if)# ip pim enable
RS G8000(config-ip-if)# ip pim component-id 2
RS G8000(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 21

Basic Redundancy

BLADEOS 6.5 includes various features for providing basic link or device redundancy:

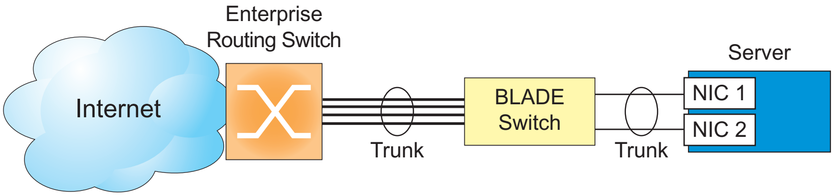
- "Trunking for Link Redundancy" on page 309
- "Hot Links" on page 310
- "Active MultiPath Protocol" on page 312
- "Stacking for High Availability Topologies" on page 316

Trunking for Link Redundancy

Multiple switch ports can be combined together to form robust, high-bandwidth trunks to other devices. Since trunks are comprised of multiple physical links, the trunk group is inherently fault tolerant. As long as one connection between the switches is available, the trunk remains active.

In Figure 34, four ports are trunked together between the switch and the enterprise routing device. Connectivity is maintained as long as one of the links remain active. The links to the server are also trunked, allowing the secondary NIC to take over in the event that the primary NIC link fails.

Figure 34 Trunking Ports for Link Redundancy



For more information on trunking, see "Ports and Trunking" on page 115.

Hot Links

For network topologies that require Spanning Tree to be turned off, 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, trunk, 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 trunk, or an LACP adminkey as a Hot Link 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, 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:

- Ports that are configured as Hot Link interfaces must have STP disabled.
- When Hot Links is turned on, MSTP, RSTP, and PVRST must be turned off.
- When Hot Links is turned on, UplinkFast must be disabled.
- 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 trunk.

Configuring Hot Links

Use the following commands to configure Hot Links.

```
RS G8000(config)# hotlinks trigger 1 enable         (Enable Hot Links Trigger 1)
RS G8000(config)# hotlinks trigger 1 master port 1 (Add port to Master interface)
RS G8000(config)# hotlinks trigger 1 backup port 2 (Add port to Backup interface)
RS G8000(config)# hotlinks enable                  (Turn on Hot Links)
```

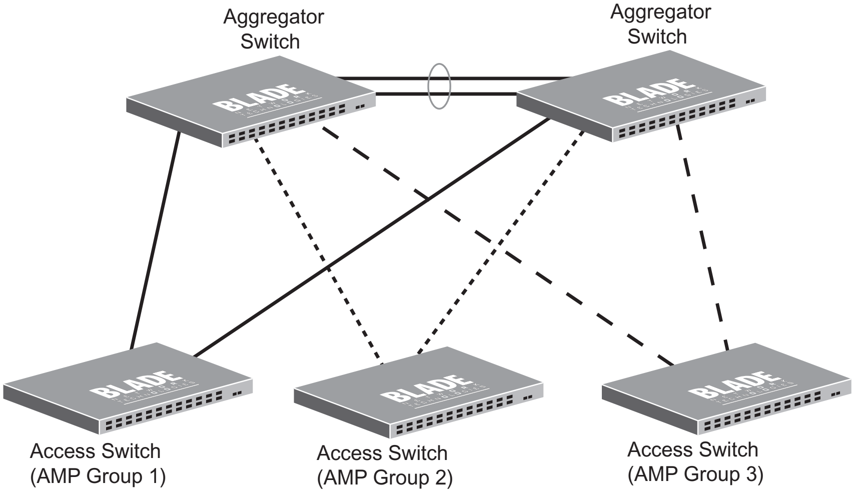
Active MultiPath Protocol

Active MultiPath Protocol (AMP) allows you to connect three switches in a loop topology, and load-balance traffic across all uplinks (no blocking). When an AMP link fails, upstream communication continues over the remaining AMP link. Once the failed AMP link re-establishes connectivity, communication resumes to its original flow pattern.

AMP is supported over Layer 2 only. Layer 3 routing is not supported. Spanning Tree is not required in an AMP Layer 2 domain. STP BPDUs will not be forwarded over the AMP links, and any BPDU packets received on AMP links are dropped.

Each AMP group contains two aggregator switches and one access switch. Aggregator switches support up to 22 AMP groups. Access switches support only one AMP group. Figure 35 shows a typical AMP topology, with two aggregators supporting a number of AMP groups.

Figure 35 AMP Topology


Aggregator Switch
Aggregator Switch
BLADE
BLADE
BLADE
BLADE
BLADE
Access Switch (AMP Group 1)
Access Switch (AMP Group 2)
Access Switch (AMP Group 3)

Each AMP group requires two links on each switch. Each AMP link consists of a single port, a static trunk group, or an LACP trunk group. Local non-AMP ports can communicate via local Layer 2 switching without passing traffic through the AMP links. No two switches in the AMP loop can have another active connection between them through a non-AMP switch.

Each AMP switch has a priority value (1-255). The switch with the lowest priority value has the highest precedence over the other switches. If there is a conflict between switch priorities, the switch with lowest MAC address has the highest precedence.

Note – For proper AMP operation, all access switches should be configured with a higher priority value (lower precedence) than the aggregators. Otherwise, some AMP control packets may be sent to access switches, even when their AMP groups are disabled.

When the AMP loop is broken, the STP port states are set to forwarding or blocking, depending on the switch priority and port/trunk precedence, as follows:

- An aggregator's port/trunk has higher precedence over an access switch's port/trunk.
- Static trunks have highest precedence, followed by LACP trunks, then physical ports.
- Between two static trunks, the trunk with the lower trunk ID has higher precedence.
- Between two LACP trunks, the trunk with the lower *admin key* has higher precedence.
- Between two ports, the port with the lowest port number has higher precedence.

Health Checks

An AMP keep-alive message is passed periodically from each switch to its neighbors in the AMP group. The keep-alive message is a BPDU-like packet that passes on an AMP link even when the link is blocked by Spanning Tree. The keep-alive message carries status information about AMP ports/trunks, and is used to verify that a physical loop exists.

An AMP link is considered healthy if the switch has received an AMP keep-alive message on that link. An AMP link is considered unhealthy if a number of consecutive AMP keep-alive messages have not been received recently on that link.

FDB Flush

When an AMP port/trunk is the blocking state, FDB flush is performed on that port/trunk. Any time there is a change in the data path for an AMP group, the FDB entries associated with the ports in the AMP group are flushed. This ensures that communication is not blocked while obsolete FDB entries are aged out.

FDB flush is performed when an AMP link goes down, and when an AMP link comes up.

Configuration Guidelines

The following configuration guidelines apply to Active MultiPath Protocol:

- The G8000 can be used as an AMP access switch or AMP aggregator switch.
- 802.1x access control must be disabled before AMP is enabled.
- Enable AMP on all switches in the AMP group before connecting the switch ports.
- Access switches should be configured with a higher priority value (lower precedence) than the aggregators. Otherwise, unexpected AMP keep-alive packets may be sent from one aggregator switches to another, even when its AMP group is disabled.
- Only one active connection (port or trunk) is allowed between switches in an AMP group.
- Spanning Tree must be disabled on AMP trunks/ports.
- Hot Links must be disabled on AMP trunk/ports.

- Private VLANs must be disabled before AMP is enabled.
- AMP ports cannot be used as monitoring ports in a port-mirroring configuration.
- Do not configure AMP ports as Layer 2 Failover control ports.
- Layer 3 routing protocols are not supported on AMP-configured switches.

Configuration Example

Configuring an Aggregator Switch

Perform the following steps to configure AMP on an aggregator switch:

1. Turn off Spanning Tree.

```
>> # spanning-tree mode disable
```

2. Define a trunk group for the aggregator-aggregator link (optional). You may use a single port-to-port link.

```
>> #  portchannel 5 port 1
>> #  portchannel 5 port 2
>> #  portchannel 5 enable
```

3. Turn AMP on, and define the aggregator.

```
>> #  active-multipath enable
>> #  active-multipath switch-type aggregator
>> #  active-multipath switch-priority 10
```

4. Configure the aggregator-aggregator link. Use the trunk group defined in step 2.

```
>> #  active-multipath aggr-portchannel 5
```

5. Define the AMP group links, and enable the AMP group.

```
>> #  active-multipath group 1 portchannel 5
>> #  active-multipath group 1 port2 10
>> #  active-multipath group 1 enable
```

Each AMP group has two links, one to each peer switch in the group. For an aggregator, one AMP group link connects to the other aggregator, and one to the access switch.

Configuring an Access Switch

Perform the following steps to configure AMP on an access switch:

1. Turn off Spanning Tree.

```
>> # spanning-tree mode disable
```

2. Turn AMP on.

```
>> # active-multipath enable
```

3. Define the AMP group links, and enable the AMP group.

```
>> # active-multipath group 1 port 3
>> # active-multipath group 1 port2 4
>> # active-multipath group 1 enable
```

Verifying AMP Operation

Display AMP group information to verify that the AMP loop is healthy.

```
>> # show active-multipath group 1 information

Group 1: enabled, topology UP
        Port 3: access
                State : forwarding
                Peer  : 00:22:00:ac:bd:00
                        aggregator, priority 10
        Port 4: aggregator
                State : forwarding
                Peer  : 00:25:03:49:82:00
                        aggregator, priority 1
```

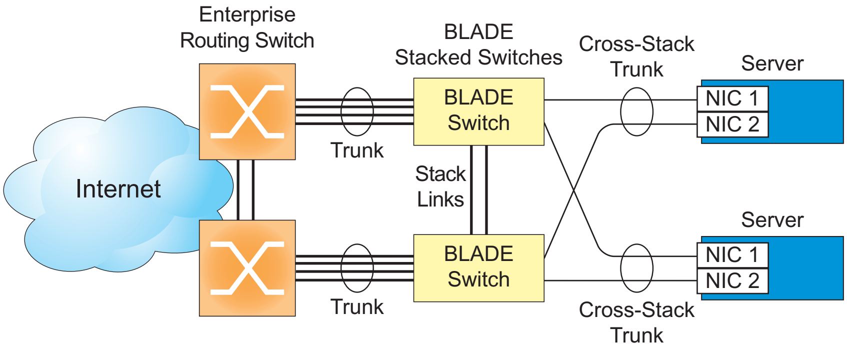
Verify that the AMP topology is `UP`, and that each link state is set to `forwarding`.

Stacking for High Availability Topologies

A *stack* is a group of up to six RackSwitch G8000 devices that work together as a unified system. Because the multiple members of a stack acts as a single switch entity with distributed resources, high-availability topologies can be more easily achieved.

In Figure 36, a simple stack using two switches provides full redundancy in the event that either switch were to fail. As shown with the servers in the example, stacking permits ports within different physical switches to be trunked together, further enhancing switch redundancy.

Figure 36 High Availability Topology Using Stacking



For more information on stacking, see "Stacking" on page 163.

CHAPTER 22

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 can be used for Layer 2 Trunk Failover (one primary and one backup). Network Adapter Teaming allows only one backup NIC for each server blade.

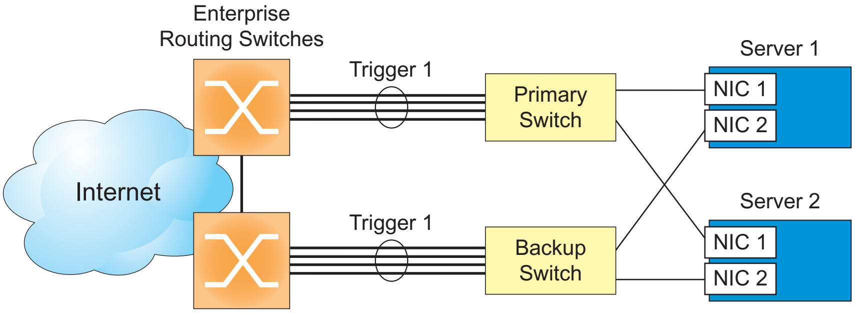
Monitoring Trunk Links

Layer 2 Failover can be enabled on any trunk group in the G8000, including LACP trunks. Trunks can be added to failover trigger groups. Then, if some specified number of monitor links fail, the switch disables all the control ports in the switch. When the control ports are disabled, it causes the NIC team on the affected servers to failover from the primary to the backup NIC. This process is called a failover event.

When the appropriate number of links in a monitor group return to service, the switch enables the control ports. This causes the NIC team on the affected servers 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 control links come up, which can take up to five seconds.

Figure 37 is a simple example of Layer 2 Failover. One G8000 is the primary, and the other is used as a backup. In this example, all ports on the primary switch belong to a single trunk group, 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 control ports. This action causes a failover event on Server 1 and Server 2.

Figure 37 Basic Layer 2 Failover



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 trunks to monitor for link failures (a monitor list), and another set of ports and/or trunks 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 trunk on the switch, or another switch.

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 trunk, 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 trunk).

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:

```
>> # show interface link                                (View port link status)
>> # show interface port <x> spanning-tree stp <x>    (View port STP status)
>> # 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 below.

LACP

Link Aggregation Control Protocol allows the switch to form dynamic trunks. You can use the *admin key* to add up to two LACP trunks to a failover trigger using automatic monitoring. When you add an *admin key* to a trigger, any LACP trunk with that *admin key* becomes a member of the trigger.

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 Listening, Learning, Blocking, or No Link). The switch automatically disables the appropriate control ports.

When the switch determines that ports in the trigger are in STP Forwarding state, then it automatically enables the appropriate control ports. The switch *fails back* to normal operation.

Configuration Guidelines

This section provides important information about configuring Layer 2 Failover.

- Any specific failover trigger can monitor ports only, static trunks only, or LACP trunks only. The different types cannot be combined in the same trigger.
- A maximum of 52 LACP keys can be added per trigger.
- Port membership for different triggers should not overlap. Any specific port should be a member of only one trigger.

Configuring Layer 2 Failover

Use the following procedure to configure a Layer 2 Failover Manual Monitor.

1. Specify the links to monitor.

```
>> # failover trigger 1 mmon monitor member 1-5
```

2. Specify the links to disable when the failover limit is reached.

```
>> # failover trigger 1 mmon control member 6-10
```

3. Configure general Failover parameters.

```
>> # failover enable
>> # failover trigger 1 enable
>> # failover trigger 1 limit 2
```

CHAPTER 23

Virtual Router Redundancy Protocol

The BNT RackSwitch G8000 (G8000) supports IPv4 high-availability network topologies through an enhanced implementation of the Virtual Router Redundancy Protocol (VRRP).

Note – BLADEOS 6.5 does not support IPv6 for VRRP.

The following topics are discussed in this chapter:

- "VRRP Overview" on page 324. This section discusses VRRP operation and BLADEOS redundancy configurations.
- "Failover Methods" on page 326. This section describes the three modes of high availability.
- "BLADEOS Extensions to VRRP" on page 328. This section describes VRRP enhancements implemented in BLADEOS.
- "Virtual Router Deployment Considerations" on page 329. This section describes issues to consider when deploying virtual routers.
- "High Availability Configurations" on page 330. This section discusses the more useful and easily deployed redundant configurations.

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 servers.

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 326 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.

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

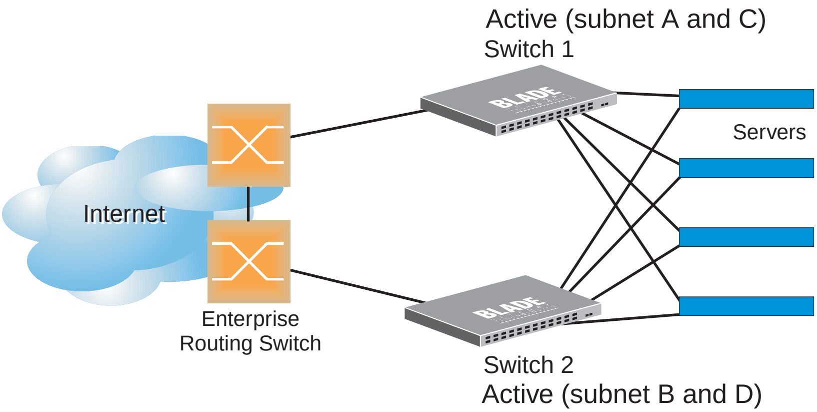
With service availability becoming a major concern on the Internet, service providers are increasingly deploying Internet traffic control devices, such as application switches, in redundant configurations. BLADEOS high availability configurations are based on VRRP. The BLADEOS implementation of VRRP includes proprietary extensions.

Active-Active Redundancy

In an active-active configuration, shown in Figure 38, 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 330.

Figure 38 Active-Active Redundancy



Virtual Router Group

The virtual router group ties all virtual routers on the switch together as a single entity. As members of a group, all virtual routers on the switch (and therefore the switch itself), are in either a master or standby state.

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.

BLADEOS Extensions to VRRP

This section describes VRRP enhancements that are implemented in BLADEOS.

BLADEOS 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.

BLADEOS can track the attributes listed in Table 19 (Router VRRP mode):

Table 19 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.
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 329 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 command to configure the virtual router ID:

```
RS G8000(config)# router vrrp
RS G8000(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 38 on page 327. 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:

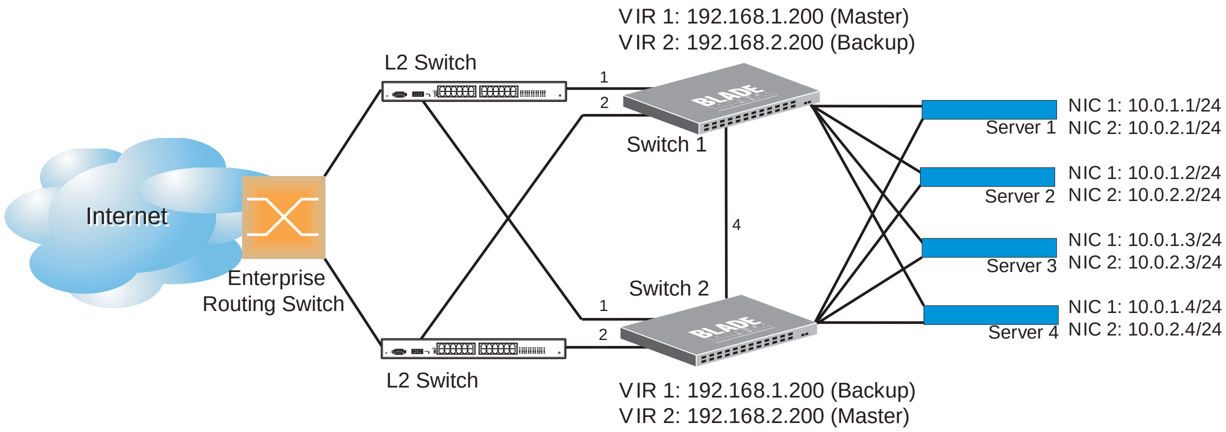
1. Set the priority for switch 1 to 101.
2. Leave the priority for switch 2 at the default value of 100.
3. On both switches, enable tracking based on ports, interfaces, or virtual routers. You can choose any combination of tracking parameters, based on your network configuration.

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

Figure 39 shows an example configuration where two G8000s are used as VRRP routers in an active-active configuration. In this configuration, both switches respond to packets.

Figure 39 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 39, traffic destined for IPv4 address 10.0.1.1 is forwarded through the Layer 2 switch at the top of the drawing, and ingresses G8000 1 on port 1. Return traffic uses default gateway 1 (192.168.1.1).

If the link between G8000 1 and the Layer 2 switch fails, G8000 2 becomes the Master because it has a higher priority. Traffic is forwarded to G8000 2, which forwards it to G8000 1 through port 4. 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 G8000 1

1. Configure client and server interfaces.

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip address 192.168.1.100 255.255.255.0
RS G8000(config-ip-if)# vlan 10
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip address 192.168.2.101 255.255.255.0
RS G8000(config-ip-if)# vlan 20
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 3
RS G8000(config-ip-if)# ip address 10.0.1.100 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 4
RS G8000(config-ip-if)# ip address 10.0.2.101 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
```

2. Configure the default gateways. Each default gateway points to a Layer 3 router.

```
RS G8000(config)# ip gateway 1 address 192.168.1.1
RS G8000(config)# ip gateway 1 enable
RS G8000(config)# ip gateway 2 address 192.168.2.1
RS G8000(config)# ip gateway 2 enable
```

3. Turn on VRRP and configure two Virtual Interface Routers.

```
RS G8000(config)# router vrrp
RS G8000(config-vrrp)# enable
RS G8000(config-vrrp)# virtual-router 1 virtual-router-id 1
RS G8000(config-vrrp)# virtual-router 1 interface 1
RS G8000(config-vrrp)# virtual-router 1 address 192.168.1.200
RS G8000(config-vrrp)# virtual-router 1 enable
RS G8000(config-vrrp)# virtual-router 2 virtual-router-id 2
RS G8000(config-vrrp)# virtual-router 2 interface 2
RS G8000(config-vrrp)# virtual-router 2 address 192.168.2.200
RS G8000(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.

```
RS G8000(config-vrrp)# virtual-router 1 track ports
RS G8000(config-vrrp)# virtual-router 1 priority 101
RS G8000(config-vrrp)# virtual-router 2 track ports
RS G8000(config-vrrp)# exit
```

5. Configure ports.

```
RS G8000(config)# vlan 10
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 1
RS G8000(config-vlan)# exit
RS G8000(config)# vlan 20
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 2
RS G8000(config-vlan)# exit
```

6. Turn off Spanning Tree Protocol globally.

```
RS G8000(config)# no spanning-tree stp 1
```

Task 2: Configure G8000 2

1. Configure client and server interfaces.

```
RS G8000(config)# interface ip 1
RS G8000(config-ip-if)# ip address 192.168.1.101 255.255.255.0
RS G8000(config-ip-if)# vlan 10
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 2
RS G8000(config-ip-if)# ip address 192.168.2.100 255.255.255.0
RS G8000(config-ip-if)# vlan 20
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 3
RS G8000(config-ip-if)# ip address 10.0.1.101 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
RS G8000(config)# interface ip 4
RS G8000(config-ip-if)# ip address 10.0.2.100 255.255.255.0
RS G8000(config-ip-if)# enable
RS G8000(config-ip-if)# exit
```

2. Configure the default gateways. Each default gateway points to a Layer 3 router.

```
RS G8000(config)# ip gateway 1 address 192.168.2.1
RS G8000(config)# ip gateway 1 enable
RS G8000(config)# ip gateway 2 address 192.168.1.1
RS G8000(config)# ip gateway 2 enable
```

3. Turn on VRRP and configure two Virtual Interface Routers.

```
RS G8000(config)# router vrrp
RS G8000(config-vrrp)# enable
RS G8000(config-vrrp)# virtual-router 1 virtual-router-id 1
RS G8000(config-vrrp)# virtual-router 1 interface 1
RS G8000(config-vrrp)# virtual-router 1 address 192.168.1.200
RS G8000(config-vrrp)# virtual-router 1 enable
RS G8000(config-vrrp)# virtual-router 2 virtual-router-id 2
RS G8000(config-vrrp)# virtual-router 2 interface 2
RS G8000(config-vrrp)# virtual-router 2 address 192.168.2.200
RS G8000(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.

```
RS G8000(config-vrrp)# virtual-router 1 track ports
RS G8000(config-vrrp)# virtual-router 2 track ports
RS G8000(config-vrrp)# virtual-router 2 priority 101
RS G8000(config-vrrp)# exit
```

5. Configure ports.

```
RS G8000(config)# vlan 10
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 1
RS G8000(config-vlan)# exit
RS G8000(config)# vlan 20
RS G8000(config-vlan)# enable
RS G8000(config-vlan)# member 2
RS G8000(config-vlan)# exit
```

6. Turn off Spanning Tree Protocol globally.

```
RS G8000(config)# no spanning-tree stp 1
```

Part 7: Network Management

CHAPTER 24

Link Layer Discovery Protocol

The BLADEOS 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 G8000 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 G8000 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.

Enabling or Disabling LLDP

Global LLDP Setting

By default, LLDP is disabled on the G8000. To turn LLDP on or off, use the following command:

```
RS G8000(config)# [no] lldp enable                (Turn LLDP on or off globally)
```

Transmit and Receive Control

The G8000 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, G8000 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:

```
RS G8000(config)# interface port 1                 (Select a switch port)
RS G8000(config-if)# lldp admin-status tx_rx       (Transmit and receive LLDP)
RS G8000(config-if)# lldp admin-status tx_only     (Only transmit LLDP)
RS G8000(config-if)# lldp admin-status rx_only     (Only receive LLDP)
RS G8000(config-if)# no lldp admin-status          (Do not participate in LLDP)
RS G8000(config-if)# exit                          (Exit port mode)
```

To view the LLDP transmit and receive status, use the following commands:

```
RS G8000(config)# show lldp port                      (status of all ports)
RS G8000(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.

Scheduled Interval

The G8000 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:

```
RS G8000(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 G8000 detects relevant changes to its configuration or status (such as when ports are enabled or disabled). To prevent the G8000 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:

```
RS G8000(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.

```
RS G8000(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 G8000 (see “Using Simple Network Management Protocol” on page 33), 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 (Interface Port mode):

```
RS G8000(config)# interface port 1
RS G8000(config-if)# [no] lldp trap-notification
RS G8000(config-if)# exit
```

In addition to sending LLDP information at scheduled intervals, LLDP information is also sent when the G8000 detects relevant changes to its configuration or status (such as when ports are enabled or disabled). To prevent the G8000 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:

```
RS G8000(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 command:

```
RS G8000(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 LLDP admin-status command options (see "Transmit and Receive Control" on page 338), 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 G8000 port from their MIB.

In addition, if LLDP is fully disabled on a port and then later re-enabled, the G8000 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:

```
RS G8000(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 338), 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 (Interface Port mode):

```
RS G8000(config)# interface port 1
RS G8000(config-if)# [no] lldp tlv <type>
RS G8000(config-if)# exit
```

where *type* is an LLDP information option from Table 20:

Table 20 LLDP Optional Information Types

Type	Description
`portdesc`	Port Description
`sysname`	System Name
`sysdescr`	System Description
`syscap`	System Capabilities
`mgmtaddr`	Management Address
`portvid`	IEEE 802.1 Port VLAN ID
`portprot`	IEEE 802.1 Port and Protocol VLAN ID
`vlanname`	IEEE 802.1 VLAN Name
`protid`	IEEE 802.1 Protocol Identity
`macphy`	IEEE 802.3 MAC/PHY Configuration/Status, including the auto-negotiation, duplex, and speed status of the port.
`powermdi`	IEEE 802.3 Power via MDI, indicating the capabilities and status of devices that require or provide power over twisted-pair copper links.
`linkaggr`	IEEE 802.3 Link Aggregation status for the port.
`framesz`	IEEE 802.3 Maximum Frame Size for the port.
`all`	Select all optional LLDP information for inclusion or exclusion.

By default, all optional LLDP information types are included in LLDP transmissions.

LLDP Receive Features

Types of Information Received

When the LLDP receive option is enabled on a port (see "Enabling or Disabling LLDP" on page 338), 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 G8000 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.

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 G8000 MIB
- Using the G8000 Browser-Based Interface (BBI)
- Using CLI or isCLI commands on the G8000

Using the CLI the following command displays remote LLDP information:

```
RS G8000(config)# show lldp remote-device [<index number>]
```

To view a summary of remote information, omit the *Index number* parameter. For example:

```
RS G8000(config)# show lldp remote-device
LLDP Remote Devices Information

LocalPort | Index | Remote Chassis ID    | Remote Port | Remote System Name
----------|-------|----------------------|-------------|---------------------
3         | 1     | 00 18 b1 33 1d 00    | 23          |
```

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:

```
RS G8000(config)# show lldp remote-device 1
Local Port Alias: 3
        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        : 7

        System Name        :
        System Description : BNT 1/10Gb Uplink Ethernet Switch Module,
                             flash image: version 5.1.0,
                             boot image: version 5.1.0.12

        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.

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.

```
RS G8000(config)# lldp enable
```

2. Set the global LLDP timer features.

```
RS G8000(config)# lldp transmission-delay 30       (Transmit each 30 seconds)
RS G8000(config)# lldp transmission-delay 2        (No more often than 2 sec.)
RS G8000(config)# lldp holdtime-multiplier 4       (Remote hold 4 intervals)
RS G8000(config)# lldp reinit-delay 2              (Wait 2 sec. after reinit.)
RS G8000(config)# lldp trap-notification-interval 5(Minimum 5 sec. between)
```

3. Set LLDP options for each port.

```
RS G8000(config)# interface port <n>          (Select a switch port)
RS G8000(config-if)# lldp admin-status tx_rx(Transmit and receive LLDP)
RS G8000(config-if)# lldp trap-notification (Enable SNMP trap notifications)
RS G8000(config-if)# lldp tlv all            (Transmit all optional information)
RS G8000(config-if)# exit
```

4. Enable syslog reporting.

```
RS G8000(config)# logging log lldp
```

5. Verify the configuration settings:

```
RS G8000(config)# show lldp
```

6. View remote device information as needed.

```
RS G8000(config)# show lldp remote-device
   or
RS G8000(config)# show lldp remote-device <index number>
```

CHAPTER 25

Simple Network Management Protocol

BLADEOS 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 or HP-OpenView.

Note – SNMP read and write functions are enabled by default. For best security practices, if SNMP is not needed for your network, it is recommended that you disable these functions prior to connecting the switch to the network.

SNMP Version 1

To access the SNMP agent on the G8000, the read and write community strings on the SNMP manager should be configured to match those on the switch. The default read community string on the switch is `public` and the default write community string is `private`.

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

```
RS G8000(config)# snmp-server read-community <1-32 characters>
    -and-
RS G8000(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:

```
RS G8000(config)# snmp-server trap-src-if <trap source IP interface>
RS G8000(config)# snmp-server host <IPv4 address> <trap host community string>
```

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 menu:

```
RS G8000(config)# snmp-server ?
```

For more information on SNMP MIBs and the commands used to configure SNMP on the switch, see the *BLADEOS 6.5 Command Reference*.

Default Configuration

BLADEOS has two SNMPv3 users by default. Both of the following users have access to all the MIBs supported by the switch:

- User 1 name is `adminmd5` (password `adminmd5`). Authentication used is MD5.
- User 2 name is `adminsha` (password `adminsha`). Authentication used is SHA.

Up to 16 SNMP users can be configured on the switch. To modify an SNMP user, enter the following commands:

```
RS G8000(config)# snmp-server user <1-16> name <1-32 characters>
```

Users can be configured to use the authentication/privacy options. The G8000 support two authentication algorithms: MD5 and SHA, as specified in the following command:

```
RS G8000(config)# snmp-server user <1-16> authentication-protocol
    {md5|sha} authentication-password

   -or-

RS G8000(config)# snmp-server user <1-16> 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 CLI commands.

```
RS G8000(config)# snmp-server user 5 name admin
RS G8000(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.

RS G8000(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.

```
RS G8000(config)# snmp-server access 5 name admingrp
RS G8000(config)# snmp-server access 5 level authpriv
RS G8000(config)# snmp-server access 5 read-view iso
RS G8000(config)# snmp-server access 5 write-view iso
RS G8000(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.

```
RS G8000(config)# snmp-server group 5 user-name admin
RS G8000(config)# snmp-server group 5 group-name admingrp
```

Configuring SNMP Trap Hosts

SNMPv1 Trap Host

1. Configure a user with no authentication and password.

```
>> # /cfg/sys/ssnmp/snmpv3/usm 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:

```
>> # /cfg/sys/ssnmp/snmpv3/access <user number>
```

In the example below the user will receive the traps sent by the switch.

```
/c/sys/ssnmp/snmpv3/access 10                  (Access group to view SNMPv1 traps)
     name "v1trap"
     model snmpv1
     nview "iso"
/c/sys/ssnmp/snmpv3/group 10                   (Assign user to the access group)
     model snmpv1
     uname v1trap
     gname v1trap
```

3. Configure an entry in the notify table.

```
RS G8000(config)# snmp-server notify 10 name v1trap
RS G8000(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:

```
RS G8000(config)# snmp-server target-address 10 name v1trap address
    10.70.70.190
RS G8000(config)# snmp-server target-address 10 parameters-name v1param
RS G8000(config)# snmp-server target-address 10 taglist v1param
RS G8000(config)# snmp-server target-parameters 10 name v1param
RS G8000(config)# snmp-server target-parameters 10 user-name v1only
RS G8000(config)# snmp-server target-parameters 10 message snmpv1
```

Note – BLADEOS 6.5 supports only IPv4 addresses for SNMP trap hosts.

5. Use the community table to specify which community string is used in the trap.

```
/c/sys/ssnmp/snmpv3/comm 10                      (Define the community string)
     index v1trap
     name public
     uname 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`.

```
RS G8000(config)# snmp-server user 10 name v2trap

RS G8000(config)# snmp-server group 10 security snmpv2
RS G8000(config)# snmp-server group 10 user-name v2trap
RS G8000(config)# snmp-server group 10 group-name v2trap
RS G8000(config)# snmp-server access 10 name v2trap
RS G8000(config)# snmp-server access 10 security snmpv2
RS G8000(config)# snmp-server access 10 notify-view iso

RS G8000(config)# snmp-server notify 10 name v2trap
RS G8000(config)# snmp-server notify 10 tag v2trap

RS G8000(config)# snmp-server target-address 10 name v2trap
    address 100.10.2.1
RS G8000(config)# snmp-server target-address 10 taglist v2trap
RS G8000(config)# snmp-server target-address 10 parameters-name
    v2param
RS G8000(config)# snmp-server target-parameters 10 name v2param
RS G8000(config)# snmp-server target-parameters 10 message snmpv2c
RS G8000(config)# snmp-server target-parameters 10 user-name v2trap
RS G8000(config)# snmp-server target-parameters 10 security snmpv2

RS G8000(config)# snmp-server community 10 index v2trap
RS G8000(config)# snmp-server community 10 user-name v2trap
```

Note – BLADEOS 6.5 supports only IPv4 addresses for SNMP 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:

```
RS G8000(config)# snmp-server access <1-32> level
RS G8000(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:

```
RS G8000(config)# snmp-server user 11 name v3trap
RS G8000(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.
RS G8000(config)# snmp-server access 11 notify-view iso
RS G8000(config)# snmp-server access 11 level authnopriv
RS G8000(config)# snmp-server group 11 user-name v3trap
RS G8000(config)# snmp-server group 11 tag v3trap
RS G8000(config)# snmp-server notify 11 name v3trap
RS G8000(config)# snmp-server notify 11 tag v3trap
RS G8000(config)# snmp-server target-address 11 name v3trap address
    47.81.25.66
RS G8000(config)# snmp-server target-address 11 taglist v3trap
RS G8000(config)# snmp-server target-address 11 parameters-name v3param
RS G8000(config)# snmp-server target-parameters 11 name v3param
RS G8000(config)# snmp-server target-parameters 11 user-name v3trap
RS G8000(config)# snmp-server target-parameters 11 level authNoPriv
```

Note – BLADEOS 6.5 supports only IPv4 addresses for SNMP trap hosts.

SNMP MIBs

The BLADEOS 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 BLADEOS SNMP agent are contained in the following BLADEOS enterprise MIB document:

The BLADEOS SNMP agent supports the following standard MIBs:

- `dot1x.mib`
- `ieee8021ab.mib`
- `ieee8023ad.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 BLADEOS 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 BLADEOS:

Table 21 BLADEOS-Supported Enterprise SNMP Traps

Trap Name	Description
`altSwDefGwUp`	Signifies that the default gateway is alive.
`altSwDefGwDown`	Signifies that the default gateway is down.
`altSwDefGwInService`	Signifies that the default gateway is up and in service
`altSwDefGwNotInService`	Signifies that the default gateway is alive but not in service
`altSwVrrpNewMaster`	Indicates that the sending agent has transitioned to “Master” state.
`altSwVrrpNewBackup`	Indicates that the sending agent has transitioned to “Backup” state.
`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.
`altSwLoginFailure`	Signifies that someone failed to enter a valid username/password combination.
`altSwTempExceedThreshold`	Signifies that the switch temperature has exceeded maximum safety limits.
`altSwTempReturnThreshold`	Signifies that the switch temperature has returned below maximum safety limits.
`altSwStgNewRoot`	Signifies that the bridge has become the new root of the STG.
`altSwStgTopologyChanged`	Signifies that there was a STG topology change.
`altSwStgBlockingState`	An `altSwStgBlockingState` trap is sent when port state is changed in blocking state.
`altSwCistNewRoot`	Signifies that the bridge has become the new root of the CIST.
`altSwCistTopologyChanged`	Signifies that there was a CIST topology change.

Table 21 BLADEOS-Supported Enterprise SNMP Traps (continued)

Trap Name	Description
`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.

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 22.

Table 22 lists the MIBS used to perform operations associated with the Switch Image and Configuration files.

Table 22 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 22.

- Load a new Switch image (boot or running) from a FTP/TFTP server
- Load a previously saved switch configuration from a FTP/TFTP server
- Save the switch configuration to a FTP/TFTP server
- Save a switch dump to a FTP/TFTP 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 server at IPv4 address 192.168.10.10, though IPv6 is also supported.

1. Set the FTP/TFTP 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 FTP server, enter a username:

   ```
   Set agTransferUserName.0 "MyName"
   ```

5. If you are using an 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 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 FTP server, enter a username:

   ```
   Set agTransferUserName.0 "MyName"
   ```

4. If you are using an 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 server follow the steps below. This example shows a FTP/TFTP server at IPv4 address 192.168.10.10, though IPv6 is also supported.

1. Set the FTP/TFTP 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 FTP server, enter a username:

   ```
   Set agTransferUserName.0 "MyName"
   ```

4. If you are using an 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 server, follow the steps below. This example shows an FTP/TFTP server at 192.168.10.10, though IPv6 is also supported.

1. Set the FTP/TFTP 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 FTP server, enter a username:

   ```
   Set agTransferUserName.0 "MyName"
   ```

4. If you are using an FTP server, enter a password:

   ```
   Set agTransferPassword.0 "MyPassword"
   ```

5. Initiate the transfer. To save a dump file, enter 5:

   ```
   Set agTransferAction.0 "5"
   ```

Part 8: Monitoring

The ability to monitor traffic passing through the G8000 can be invaluable for troubleshooting some types of networking problems. This sections cover the following monitoring features:

- Remote Monitoring (RMON)
- sFLOW
- Port Mirroring

CHAPTER 26

Remote Monitoring

Remote Monitoring (RMON) allows network devices to exchange network monitoring data.

RMON allows the switch to perform the following functions:

- Track events and trigger alarms when a threshold is reached.
- Notify administrators by issuing a syslog message or SNMP trap.

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:

- Group 1: Statistics
- Group 2: History
- Group 3: Alarms
- 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. You can configure RMON statistics on a per-port basis.

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.

Example Configuration

1. Enable RMON on a port.

```
RS G8000(config)# interface port 1
RS G8000(config-if)# rmon
```

2. View RMON statistics for the port.

```
RS G8000(config-if)# show interface port 1 rmon-counters
------------------------------------------------------------------
RMON statistics for port 3:
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. History sampling is done per port.

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 index 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.

You can use an SNMP browser to view History samples.

History MIB Object ID

The type of data that can be sampled must be of an `ifIndex` object type, as described in RFC 1213 and RFC 1573. The most common data type for the History sample is as follows:

```
1.3.6.1.2.1.2.2.1.1.<x>
```

The last digit (*x*) represents the number of the port to monitor.

Configuring RMON History

Perform the following steps to configure RMON History on a port.

1. Enable RMON on a port.

```
RS G8000(config)# interface port 1
RS G8000(config-if)# rmon
RS G8000(config-if)# exit
```

2. Configure the RMON History parameters for a port.

```
RS G8000(config)# rmon history 1 interface-oid 1.3.6.1.2.1.2.2.1.1.<x>
RS G8000(config)# rmon history 1 requested-buckets 30
RS G8000(config)# rmon history 1 polling-interval 120
RS G8000(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.

```
RS G8000(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:

```
RS G8000(config)# rmon alarm <alarm number> rising-crossing-index
      <event number>
RS G8000(config)# rmon alarm <alarm number> falling-crossing-index
      <event number>
```

When the alarm threshold is reached, the corresponding event is triggered.

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 statistic follows:

```
1.3.6.1.2.1.5.1.0 – mgmt.icmp.icmpInMsgs
1.3.6.1.2.1.2.2.1.10.x – ifInOctets
```

The last digit (*x*) represents the interface on which to monitor, which corresponds to the interface number, or port number, as follows:

```
1-128 = IF 1-128
129 = port 1
130 = port 2
…
172 = port 44
```

This value represents the alarm’s MIB OID, as a string. Note that for non-tables, you must supply a `.0` to specify end node.

Configuring RMON Alarms

Example 1

Configure the RMON Alarm parameters to track the number of packets received on port 1.

```
RS G8000(config)# rmon alarm 1 oid 1.3.6.1.2.1.2.2.1.10.129
RS G8000(config)# rmon alarm 1 alarm-type rising
RS G8000(config)# rmon alarm 1 rising-crossing-index 100
RS G8000(config)# rmon alarm 1 interval 3600
RS G8000(config)# rmon alarm 1 rising-limit 2000000000
RS G8000(config)# rmon alarm 1 owner "Alarm for ifInOctets"
```

This configuration creates an RMON alarm that checks `ifInOctets` on port 1 once every hour. If the statistic exceeds two billion, an alarm is generated that triggers event index 100.

Example 2

Configure the RMON Alarm parameters to track ICMP messages.

```
RS G8000(config)# rmon alarm 1 oid 1.3.6.1.2.1.5.8.0
RS G8000(config)# rmon alarm 1 alarm-type rising
RS G8000(config)# rmon alarm 1 rising-crossing-index 110
RS G8000(config)# rmon alarm 1 interval-time 60
RS G8000(config)# rmon alarm 1 rising-limit 200
RS G8000(config)# rmon alarm 1 sample delta
RS G8000(config)# rmon alarm 1 owner "Alarm for icmpInEchos"
```

This 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 110.

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, 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:

```
RS G8000(config)# rmon alarm <alarm number> rising-crossing-index
        <event number>
RS G8000(config)# rmon alarm <alarm number> falling-crossing-index
        <event number>
```

RMON events use SNMP and syslogs 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 syslog of type RMON that corresponds to the event.

For example, to configure the RMON event parameters.

```
RS G8000(config)# rmon event 110 type log
RS G8000(config)# rmon event 110 description "SYSLOG_this_alarm"
RS G8000(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 27
sFLOW

The G8000 supports sFlow technology for monitoring traffic in data networks. The switch includes an embedded sFlow agent which can be configured to 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 – BLADEOS 6.5 does not support IPv6 for sFLOW.

---

sFlow Statistical Counters

The G8000 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 G8000 reports general port statistics and port Ethernet statistics.

sFlow Network Sampling

In addition to statistical counters, the G8000 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):

```
RS G8000(config)# sflow server <IPv4 address>      (sFlow server address)
RS G8000(config)# sflow port <service port>        (Set the optional service port)
RS G8000(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 `no sflow enable` command.

2. On a per-port basis, define the statistics polling rate:

```
RS G8000(config)# interface port <port>
RS G8000(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:

```
RS G8000(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.

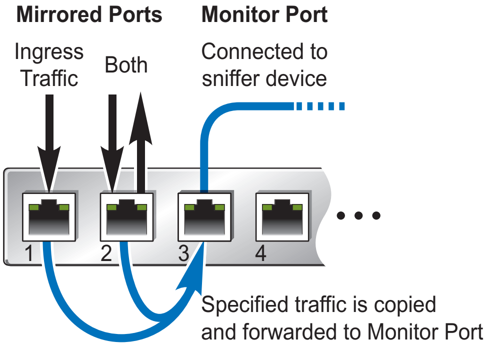
CHAPTER 28

Port Mirroring

The BLADEOS 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 G8000 supports a "many to one" mirroring model. As shown in Figure 40, selected traffic for ports 1 and 2 is being monitored by port 3. In the example, both ingress traffic and egress traffic on port 2 are copied and forwarded to the monitor. However, port 1 mirroring is configured so that only ingress traffic is copied and forwarded to the monitor. A device attached to port 3 can analyze the resulting mirrored traffic.

Figure 40 Mirroring Ports


Mirrored Ports
Monitor Port
Ingress Traffic
Both
Connected to sniffer device
1
2
3
4
Specified traffic is copied and forwarded to Monitor Port

The G8000 supports three monitor ports in stand-alone (non-stacking) mode. Only one monitor port is supported in stacking mode. Each monitor port can receive mirrored traffic from any number of target ports.

BLADEOS 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 1 traffic cannot be monitored by both port 3 and 4 at the same time, nor can port 2 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.

Configuring Port Mirroring

The following procedure may be used to configure port mirroring for the example shown in Figure 40 on page 371:

1. Specify the monitoring port, the mirroring port(s), and the port-mirror direction.

```
RS G8000(config)# port-mirroring monitor-port 3 mirroring-port 1 in
RS G8000(config)# port-mirroring monitor-port 3 mirroring-port 2 both
```

2. Enable port mirroring.

```
RS G8000(config)# port-mirroring enable
```

3. View the current configuration.

```
RS G8000# show port-mirroring

Port Monitoring : Enabled

Monitoring Ports     Mirrored Ports
1                    none
2                    none
3                    (1, in) (2, both)
4                    none
5                    none
6                    none
7                    none
8                    none
9                    none
10                   none
...
```

Part 9: Appendices

APPENDIX A
Glossary

CNA	Converged Network Adapter. A device used for I/O consolidation such as that in Converged Enhanced Ethernet (CEE) environments implementing Fibre Channel over Ethernet (FCoE). The CNA performs the duties of both a Network Interface Card (NIC) for Local Area Networks (LANs) and a Host Bus Adapter (HBA) for Storage Area Networks (SANs).
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 G8000s 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.

Index