Chapter 21

# Configuring QoS for 802.3ad Link Aggregation Groups

This chapter provides information for configuring QoS for 802.3ad link aggregation groups.

QoS topics are discussed in the following sections:

## QoS for 802.3ad Link Aggregation Interfaces Overview

You can configure QoS for 802.3ad link aggregation interfaces. To ensure that QoS is applied properly to the interface column, you configure the QoS profile using either a hashed load-balancing scheme or a subscriber load-balancing scheme.

### *Types of Load Balancing*

For hashed load balancing, you configure the scheduler hierarchy with Ethernet queues, and the system replicates them on each link within the link aggregation group (LAG). The system demultiplexes each packet to one of the active links in the LAG using a random hash generated by fields in the packet header. For example, when an IP packet is routed to a LAG, the hash algorithm is based on the IP Source Address and Destination Address in the IP header.

For subscriber load balancing, you configure the scheduler hierarchy with IP, VLAN, and S-VLAN queues and the system allocates them to individual ports in the LAG. The system demultiplexes each packet to an active link based on the subinterface underlying the egress interface. For example, when an IP packet is routed to an IP interface over a LAG, the system binds the underlying VLAN, PPPoE, or MPLS subinterface to one of the active links in the LAG. The packet is transmitted over the interface.

Most network operators configure QoS over 802.3ad LAGs using subscriber load balancing to take advantage of subscriber class-based queueing (SCBQ) features. However, configuring hashed load balancing is useful for achieving fine-grained distribution of multicast VLAN traffic or for any high bandwidth VLAN that does not require shared shaping.

To ensure that QoS is symmetrically applied to all the links, the router periodically rebalances the traffic on the LAG. You can control the load-balancing parameters.

### *Munged QoS Profiles and Load Balancing*

To determine whether to use hashed load balancing or subscriber load balancing, the system munges a QoS profile for a subscriber.

In typical Ethernet configurations, the munged QoS profile for a given subscriber interface comprises the accumulated rules of the QoS profiles attached below the subscriber interface in the interface column. Rules in higher-attached QoS profiles override or eclipse rules in lower-attached QoS profiles. For example, rules from specific interface attachments such as a VLAN override those from attachments at S-VLANs or ports.

When applying QoS to LAGs, the system uses a modified algorithm to munge QoS profile attachments. The system automatically builds the munged QoS profile using the rules in the QoS profile attached at the LAG interface.

For example, the munged Qos profile for VLAN 0,0 consists of the munge of:

- Attachment 1—QoS profile attached to the VLAN
- Attachment 2—QoS profile attached to the S-VLAN
- Attachment 3—QoS profile attached to the LAG

If there is no QoS profile attached to the LAG, the system locates the lag-default QoS profile indicated in the **qos-port-type-profile** command.

If the resulting QoS profile specifies only Ethernet queues, the system uses the hash algorithm to balance the links. If the resulting QoS profile specifies any VLAN, IP, or L2TP-Session queues, then the system uses subscriber load balancing.

## *802.3ad Link Aggregation and QoS Parameters*

You can create parameter instances for IEEE 803.ad LAG interfaces. A parameter instance for LAG can control an Ethernet port or a node, but you cannot create parameter instances for the Ethernet interfaces within the LAG.

For example, a LAG instance can specify a shaping rate of 100 Mbps on an Ethernet port or a group node. The system shapes all Ethernet ports or group nodes to the same rate within the LAG. Using load balancing, the system strives to balance the traffic each link equally.

## *QoS and Ethernet Link Redundancy*

You can configure Ethernet link redundancy for LAG interfaces. When you configure QoS for those links, be sure to consider the following behaviors.

### Active Link Failure and QoS

When an active link fails, traffic that is hashed-load balanced is redirected onto the remaining active links in the LAG. Traffic that is hashed-load balanced might be lost on the disabled link, but from the moment of switchover, traffic arriving from the fabric on the egress line module is directed towards one of the remaining hashed load-balanced queues.

Subscriber load-balanced traffic takes more time to reestablish on active links because of the amount of computation (approximately 3 ms per subscriber). During this time period, traffic directed to the disabled link might be lost.

### Administratively Disabling a Link and QoS

When a link is administratively disabled, the system immediately redirects traffic from the link to other links in the LAG.

### Adding a New Link to the LAG and QoS

When you add a new link to the LAG, the system immediately sends traffic that is hashed-load balanced to the link. Traffic that is subscriber-load balanced moves to the new link as new subscribers log in. The system automatically rebalances traffic to the new link based on the load rebalance configuration for the LAG.

## *Related Topics*

- Hashed Load Balancing for 802.3ad Link Aggregation Groups Overview on page 198
- Subscriber Load Balancing for 802.3ad Link Aggregation Groups Overview on page 199

- For more information about configuring the lag-default QoS profile for default subscriber load balancing, see *Enabling Default Subscriber Load Balancing for 802.3ad Link Aggregation Groups* on page 204
- For more information about Ethernet link redundancy, see *JUNOSe Link Layer Configuration Guide, Chapter 6, Configuring 802.3ad Link Aggregation and Link Redundancy*
- For more information about configuring QoS parameters, see *Configuring a QoS Parameter* on page 225
- For more information about the munge algorithm, see *Munged QoS Profile Overview* on page 142
- For a list of modules that support 802.3ad link aggregation, see the *ERX Module Guide* and the *E120 and E320 Module Guide*

## Hashed Load Balancing for 802.3ad Link Aggregation Groups Overview

To configure hashed load balancing, you configure a scheduler hierarchy with Ethernet queues and the system replicates the queues for each link within the LAG. The system shares the traffic equally across the links based on the distribution characteristics defined in the hash algorithm.

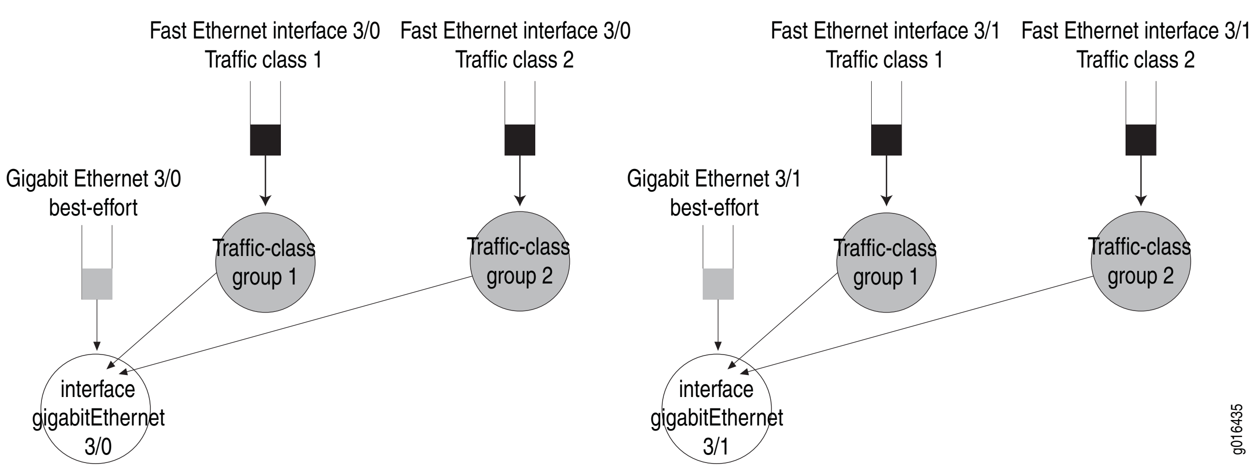
Because all traffic is carried in Ethernet queues, per-subscriber QoS features such as shared shaping for VLANs are not available.

### Sample Scheduler Hierarchy for Hashed Load Balancing

Figure 48 displays a sample 802.3ad link aggregation scheduler hierarchy that uses hashed load balancing.

The Gigabit Ethernet interfaces are on the same line module and are members of a LAG. The system dynamically balances traffic between the Ethernet queues on the two ports.

**Figure 48: 802.3ad Link Aggregation Scheduler Hierarchy**



## Subscriber Load Balancing for 802.3ad Link Aggregation Groups Overview

To configure subscriber load balancing, you configure a scheduler hierarchy with nodes and queues for IP, VLANs, and S-VLANs. The system distributes those nodes and queues in the scheduler hierarchy over the ports within the LAG using a technique called *partitioning*.

Ethernet queues used for hashed load balancing are always present in the scheduler hierarchy.

To ensure that QoS is symmetrically applied to all the links, the router periodically rebalances the load within the LAG using a hash algorithm. You can control the load-balancing parameters and configure the system to dynamically rebalance. Partitioning the Scheduler Hierarchy

The system then partitions the scheduler hierarchy by binding the IP, VLAN, L2TP session, and MPLS resources for each subscriber to a selected link within the LAG at the time the subscriber interface is configured.

### S-VLANs and Subscriber Load Balancing

The system *clones* S-VLAN nodes and queues on each link in the LAG. The system clone S-VLANs so it can allocate subscribers that share a common S-VLAN ID to different links within the LAG. S-VLAN nodes and queues are the only resources that are cloned; the system always allocates nodes and queues for other interface types to a single selected link.

Cloning S-VLAN nodes enables fine-grained load balancing within the LAG because VLANs within the S-VLAN can be allocated to the link with the least traffic. However, cloned S-VLANs can introduce anomalous scheduling behavior. A shaped S-VLAN node within the LAG shapes traffic on a per-link basis. Shaping a LAG S-VLAN node to 2 Mbps on a LAG with 2 links can enable up to 4 Mbps of traffic (2 Mbps per link).

Shared shaping on an S-VLAN within a LAG has the same behavior; the LAG S-VLAN that is shared shaped to 10 Mbps on a LAG with 2 ports allows up to 20Mbps of traffic; 10 Mbps for each link.

### PPPoE over VLANs and Subscriber Load Balancing

The system binds PPPoE subscribers stacked over a common VLAN to the same link within the LAG. Because the underlying VLAN node is allocated to a single link, the system allocates all traffic over that VLAN to that link.

### PPPoE over Ethernet (No VLANs) and Subscriber Load Balancing

The system allocates subscribers to each link independently. There are no S-VLAN nodes to clone, and no related VLAN nodes that require allocation on the same link.

### MPLS over LAG and Subscriber Load Balancing

For QoS purposes, the system considers base tunnels as logical interfaces, but does not consider stacked tunnels. The system assigns MPLS base tunnels stacked over VLANs to the link to which the VLAN is assigned.

## Sample Scheduler Hierarchy for Subscriber Load Balancing

Figure 49 on page 200 displays the scheduler hierarchy for the Gigabit Ethernet interface in slot 3, port 0. Figure 50 on page 201 displays the scheduler hierarchy for the Gigabit Ethernet interface in slot 3, port 1.

The Ethernet queues are shown in gray; they are not bound to a link in the LAG and are replicated for each link in the LAG. These queues are used for subscribers with QoS profiles that indicate Ethernet queues, and for traffic classes other than best-effort, traffic class 1, and traffic class 2.

When partitioning the scheduler hierarchy that includes 1000 VLAN subinterfaces, the system binds 500 of the subinterfaces to port 0, and binds another 500 to port 1. The binding for a given VLAN subinterface is arbitrary.

The scheduler nodes for a given VLAN subinterface are always allocated to the same port within the LAG. In this example, the scheduler nodes for VLAN 0,0 are all allocated to Gigabit Ethernet interface in slot 3, port 0.

S-VLAN nodes and queues are cloned for each link in the LAG. S-VLAN nodes in each traffic-class group are shown identically on both ports.

**Figure 49: Subscriber Load-Balanced Scheduler Hierarchy for Port 0**

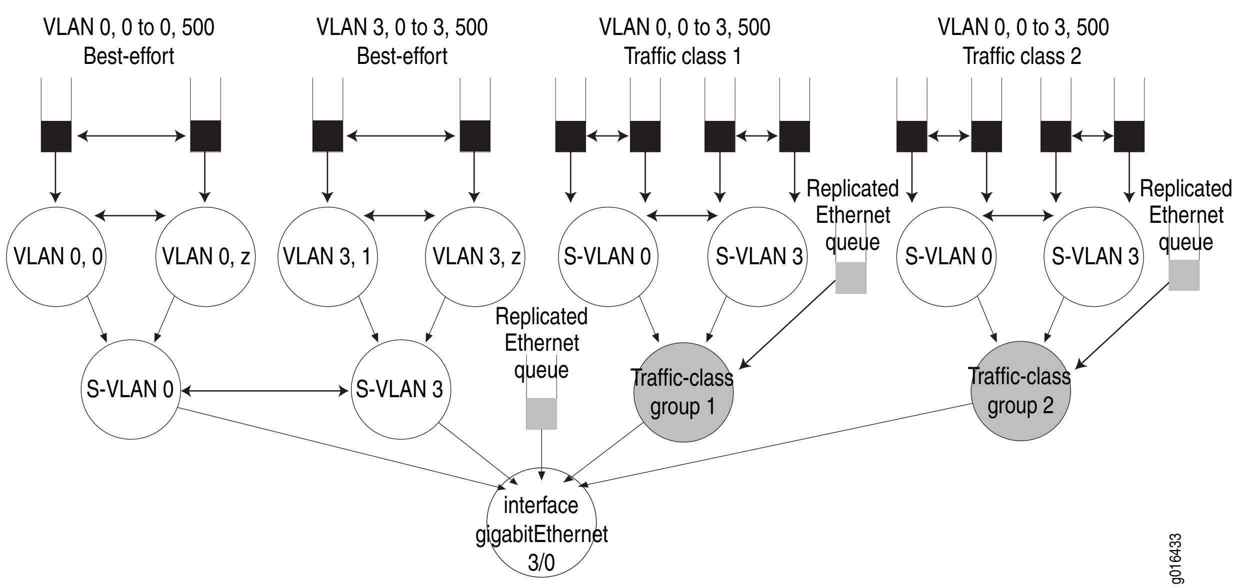
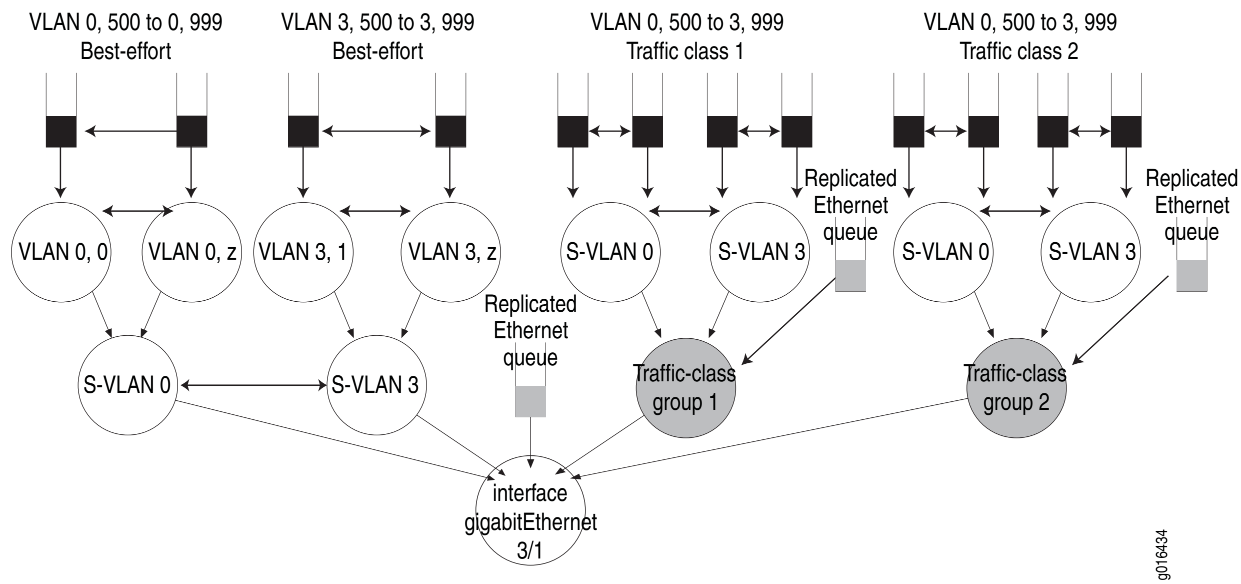


**Figure 50: Subscriber Load-Balanced Scheduler Hierarchy for Port 1**



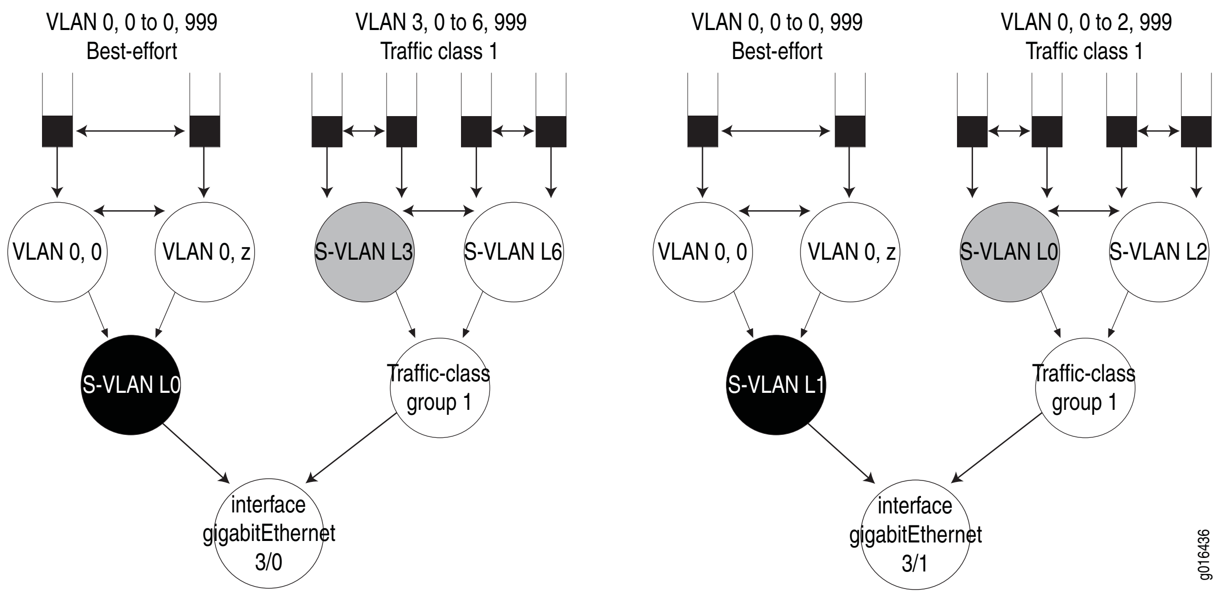

## *Subscriber Allocation in 802.3ad Link Aggregation Groups*

You can configure upper-layer subinterfaces over a LAG interface, including VLANs, PPPoE, and MPLS.

The system balances any upper-layer subinterfaces so that each active link in the LAG carriers an equal number of upper-layer subinterfaces. For this purpose, the system counts each upper-layer interface as a single subscriber, regardless of the number of forwarding interfaces stacked above it.

Figure 51 on page 202 displays a sample allocation of subscribers.

**Figure 51: Subscriber Allocation and Load Balancing**




In an ideal QoS configuration, queues and nodes are stacked over a single port that corresponds to a LAG, with the port bandwidth equal to the sum of the overall port bandwidth.

However, the actual LAG behavior is different. No level 1 node or queue can exceed the bandwidth of a link. The relative weighting of queues and nodes results in proportional bandwidth allocation only within a link, but not across the entire LAG. Actual traffic might not be evenly balanced across links in the LAG, resulting in latency and loss on one link, while another link may be lightly loaded.

Even though relative weighting is different on a LAG, shaping and shared shaping in the partitioned scheduler hierarchy operate in the same way as a typical Ethernet configuration.

### *Related Topics*

- For more information about load rebalancing, see *Configuring Load Rebalancing for 802.3ad Link Aggregation Groups* on page 205

## Guidelines for Configuring QoS over 802.3ad Link Aggregation Groups

When you configure QoS over 802.3ad LAGs, be sure to consider the following behaviors:

- QoS profiles cannot be attached to Ethernet ports if the port is a member of a LAG. In typical QoS configurations, the Ethernet interface is considered the root of the interface hierarchy. When you configure QoS for 802.3ad link aggregation, the LAG interface is considered the root of the interface hierarchy.

- You cannot configure hierarchical QoS for IP configured directly over a LAG interface.

- You cannot obtain QoS information or statistics for IP interfaces stacked over a LAG interface using any of the **show** commands for QoS. Instead, the **show qos scheduler-hierarchy** command is designed to find the interface hierarchy rooted at the specified interface and report all scheduler nodes and queues managed by those interfaces. The typical defaults in QoS profiles such as ethernet-default and atm-default specify the "ip queue traffic-class best-effort" rule, so those queues are reported in the interface hierarchy. The lag-default QoS profile does not specify this rule by default.

- Do not attach QoS profiles to IP or VLAN subinterfaces in a LAG that contain downreferences (that is, rules for S-VLAN or Ethernet nodes or queues). QoS profiles attached at subinterfaces above a LAG that also include downreference create an asymmetric scheduler hierarchy. For example, one Ethernet port might be shaped and not another.

  Also, if the QoS profile specifies only Ethernet, then the traffic sent to the subinterface might be only partially affected by the QoS profile, or not at all. The traffic can be allocated to another port entirely.

### *Related Topics*


- Subscriber Load Balancing for 802.3ad Link Aggregation Groups Overview on page 199

- Configuring the Scheduler Hierarchy for Subscriber Load Balancing in 802.3ad Link Aggregation Groups on page 204

- Configuring Load Rebalancing for 802.3ad Link Aggregation Groups on page 205


## Configuring the Scheduler Hierarchy for Hashed Load Balancing in 802.3ad Link Aggregation Groups

The type of load balancing that the system performs depends on the configuration of the scheduler hierarchy in the QoS profile.

To configure the scheduler hierarchy for hashed load balancing:

1. Configure a QoS profile.

   host1(config)#**qos-profile hashed-lag**

2. Configure the nodes and queues, including an Ethernet queue.

   host1(config-qos-profile)#**ethernet queue traffic-class best-effort**
   host1(config-qos-profile)#**ethernet queue traffic-class tc1**
   host1(config-qos-profile)#**ethernet queue traffic-class tc2**

3. Create the LAG interface and attach the QoS profile.

   host1(config)#**interface lag lg1**
   host1(config-if)#**qos-profile hashed-lag**

### Related Topics

- QoS for 802.3ad Link Aggregation Interfaces Overview on page 196
- Hashed Load Balancing for 802.3ad Link Aggregation Groups Overview on page 198
- **interface lag** command
- **node** command
- **qos-profile** command
- **queue** command

## Enabling Default Subscriber Load Balancing for 802.3ad Link Aggregation Groups

The factory default contents of the lag-default QoS profile include an Ethernet queue and the best-effort traffic class.

When you use the lag-default QoS profile, the system automatically sends traffic to the Ethernet queue and uses hash load balancing for the Ethernet queues.

To enable subscriber load balancing as the default behavior for all LAGs, issue the following command:

host1(config)#**qos-port-type-profile lag qos-profile ethernet-default**

### Related Topics

- QoS for 802.3ad Link Aggregation Interfaces Overview on page 196
- **qos-port-type-profile** command

## Configuring the Scheduler Hierarchy for Subscriber Load Balancing in 802.3ad Link Aggregation Groups

The type of load balancing that the system performs depends on the configuration of the scheduler hierarchy in the QoS profile.

To configure the scheduler hierarchy for subscriber load balancing:

1. Configure the QoS profile.

   host1(config)#**qos-profile subscriber-lag**

2. Configure the queues and nodes for VLANs and S-VLANs.

   host1(config-qos-profile)#**vlan queue traffic-class best-effort**
   host1(config-qos-profile)#**vlan queue traffic-class tc1**
   host1(config-qos-profile)#**vlan node scheduler-profile subscriber**
   host1(config-qos-profile)#**svlan node scheduler-profile svlan**
   host1(config-qos-profile)#**svlan node group g1 scheduler-profile svlan**

3. Create the LAG interface and assign member interfaces.

   host1(config)#**interface lag lg1**
   host1(config-if)#**member-interface gigabitEthernet 3/0**
   host1(config-if)#**member-interface gigabitEthernet 3/1**

4. Attach the QoS profile to the LAG interface.

   host1(config-if)#**qos-profile subscriber-lag**

### *Related Topics*

- QoS for 802.3ad Link Aggregation Interfaces Overview on page 196
- Subscriber Load Balancing for 802.3ad Link Aggregation Groups Overview on page 199
- Enabling Default Subscriber Load Balancing for 802.3ad Link Aggregation Groups on page 204
- **interface lag** command
- **member-interface** command
- **node** command
- **qos-profile** command
- **queue** command

## Configuring Load Rebalancing for 802.3ad Link Aggregation Groups

You can configure the parameters that the system uses to rebalance the links in a LAG. You can also configure the system to dynamically rebalance the links in the LAG.

Tasks to configure load rebalancing are:

- Configuring Load–Rebalancing Parameters on page 205
- Configuring the System to Dynamically Rebalance the LAG on page 207

### *Configuring Load–Rebalancing Parameters*

To configure load–rebalancing parameters:

1. Specify the LAG interface.

   host1(config)#**interface lag lg1**

2. Configure parameters that guide the system to rebalance.

   host1(config-if)#**load-rebalance period 120 start-threshold 20 percent stop-threshold 100 percent maximum-improvement 300**

This example specifies that the system rebalance within 120 seconds, can accept imbalance in the LAG in the range 20–100 percent, and can move 300 subscribers to other ports during that time.

Table 21 describes the load balancing algorithm parameters that you can configure.

**Table 21: Load Balancing Algorithm Parameters**

| Keyword | Description |
|---|---|
| **period** | Specifies the time period for rebalancing. For example, a period of 120 specifies that rebalancing occurs once every 2 minutes. |
| **start-threshold** | Specifies the amount of imbalance in the LAG that triggers the algorithm to start rebalancing. The default is 0 percent. Optionally, you can specify one of the following units of measure:<br>■ **percent**—Specifies that the amount of imbalance is measured as a percentage of the average load per link. The range is 0–100 percent.<br>For example, the average load per link in a LAG is 500. Specifying **start-threshold 5 percent** indicates that the algorithm rebalances any link that deviates from the average load per link by 25 (5 percent of 500).<br>■ **subscribers**—Specifies that the amount of imbalance is measured by the number of subscribers from the average subscriber count in the LAG. The range is 0–10000.<br>For example, specifying **start-threshold 20 subscribers** indicates that the algorithm rebalances any link with a subscriber count that differs from the average subscriber count by more than 20. |
| **stop-threshold** | Specifies the amount of imbalance in the LAG that triggers the algorithm to stop rebalancing. The algorithm continues rebalancing until this value is reached. The default is 0 percent. Optionally, you can specify one of the following units of measure:<br>■ **percent**—Specifies that the amount of imbalance is measured as a percentage of the average load per link. The range is 0–100 percent.<br>For example, the average load per link in a LAG is 500. Specifying the **stop-threshold 2 percent** command indicates that the algorithm stops within 10 of 500 (2 percent of 500). In this case, the algorithm stops when the links are at 510 and 490.<br>■ **subscribers**—Specifies that the amount of imbalance is measured by the number of subscribers. The range is 0–10000.<br>For example, specifying **stop-threshold 100 subscribers** indicates that the algorithm continues until each link in the LAG is within 100 subscribers of the average subscriber count. |
| **maximum-improvement** | Specifies the maximum number of links to rebalance in the LAG per period. The default is 100 percent. Optionally, you can specify one of the following units of measure:<br>■ **percent**—Specifies that the maximum number of links is measured as a percentage of the total links. The range is 0–100 percent.<br>For example, specifying **maximum-improvement 1 percent** indicates that the algorithm rebalances 10 links per period (1 percent of 1000).<br>■ **subscribers**—Specifies that the maximum number of links is measured by the number of subscribers. The range is 0–10000 subscribers.<br>For example, specifying **maximum-improvement 40 subscribers** indicates that the algorithm rebalances 40 subscribers per period. |

### *Configuring the System to Dynamically Rebalance the LAG*

To configure the system to dynamically rebalance the LAG:

1. Specify the LAG interface.

   host1(config)#**interface lag lg1**

2. Issue the load balance command with no keywords:

   host1(config-if)#**load-rebalance**

### *Related Topics*

- Configuring Load Rebalancing for 802.3ad Link Aggregation Groups on page 205
- QoS for 802.3ad Link Aggregation Interfaces Overview on page 196
- Enabling Default Subscriber Load Balancing for 802.3ad Link Aggregation Groups on page 204
- Configuring the Scheduler Hierarchy for Subscriber Load Balancing in 802.3ad Link Aggregation Groups on page 204
- **interface lag** command
- **load-rebalance** command

## Monitoring QoS Configurations for 802.3ad Link Aggregation Groups

To monitor Ethernet configurations for QoS:

- Monitoring the QoS Configuration of Fast Ethernet, Gigabit Ethernet, and 10-Gigabit Ethernet Interfaces on page 342
- Monitoring the QoS Configuration of IEEE 802.3ad Link Aggregation Group Bundles on page 344
- Monitoring the QoS Configuration of IP Interfaces on page 341
- Monitoring the QoS Profiles Attached to an Interface on page 335
- Monitoring the Configuration of QoS Port-Type Profiles on page 337
- Monitoring the Configuration of QoS Profiles on page 337
- Monitoring the QoS Scheduler Hierarchy on page 322
- Monitoring Shared Shapers on page 327