

Modular QoS Configuration Guide for Cisco NCS 540 Series Routers, Cisco IOS XR Release 26.1.x, 26.2.x, 26.3.x

First Published: 2026-02-28

Last Modified: 2026-09-11

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

THE SPECIFICATIONS AND INFORMATION REGARDING THE PRODUCTS IN THIS MANUAL ARE SUBJECT TO CHANGE WITHOUT NOTICE. ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS MANUAL ARE BELIEVED TO BE ACCURATE BUT ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. USERS MUST TAKE FULL RESPONSIBILITY FOR THEIR APPLICATION OF ANY PRODUCTS.

THE SOFTWARE LICENSE AND LIMITED WARRANTY FOR THE ACCOMPANYING PRODUCT ARE SET FORTH IN THE INFORMATION PACKET THAT SHIPPED WITH THE PRODUCT AND ARE INCORPORATED HEREIN BY THIS REFERENCE. IF YOU ARE UNABLE TO LOCATE THE SOFTWARE LICENSE OR LIMITED WARRANTY, CONTACT YOUR CISCO REPRESENTATIVE FOR A COPY.

The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB's public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

NOTWITHSTANDING ANY OTHER WARRANTY HEREIN, ALL DOCUMENT FILES AND SOFTWARE OF THESE SUPPLIERS ARE PROVIDED “AS IS" WITH ALL FAULTS. CISCO AND THE ABOVE-NAMED SUPPLIERS DISCLAIM ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

All printed copies and duplicate soft copies of this document are considered uncontrolled. See the current online version for the latest version.

Cisco has more than 200 offices worldwide. Addresses and phone numbers are listed on the Cisco website at www.cisco.com/go/offices.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: https://www.cisco.com/c/en/us/about/legal/trademarks.html. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1721R)

© 2026 Cisco Systems, Inc. All rights reserved.



CONTENTS

CHAPTER 1 **YANG Data Models for QoS Features** 1

Using YANG Data Models 1

CHAPTER 2 **Key Hardware Resources for QoS** 3

View Packet Processing and Traffic Management Resources 5

CHAPTER 3 **Configuring Modular QoS Service Packet Classification** 11

Packet Classification Overview 11

Traffic Class Elements 12

Default Traffic Class 12

Create a Traffic Class 12

Traffic Policy Elements 14

Create a Traffic Policy 15

Attach a Traffic Policy to an Interface 16

Packet Marking 18

Handling QoS for Locally Originated Packets 21

QoS packet marking across network layers 25

Protecting Locally Originated BFD Packets 31

Bundle Traffic Policies 32

In-Place Policy Modification 32

References for Modular QoS Service Packet Classification 33

Specification of the CoS for a Packet with IP Precedence 33

IP Precedence Bits Used to Classify Packets 34

IP Precedence Value Settings 34

IP Precedence Compared to IP DSCP Marking 34

CHAPTER 4 **Configuring Modular QoS Congestion Avoidance** 35

Modular QoS Congestion Avoidance 35

Tail Drop and the FIFO Queue 36

Configure Tail Drop 36

Random Early Detection and TCP 37

Configure Random Early Detection 38

Weighted Random Early Detection 40

Average Queue Size for WRED 40

Configure Weighted Random Early Detection 41

CHAPTER 5 **Configuring Modular QoS Congestion Management** 45

Congestion Management Overview 45

Best practices for minimizing packet drops during sustained maximum line-rate traffic 46

Class-based Weighted Fair Queueing 46

Bandwidth Remaining 46

Configuring Bandwidth Remaining 46

Low-Latency Queuing with Strict Priority Queuing 49

Configuring Low Latency Queuing with Strict Priority queuing 49

Overhead Accounting 52

Ingress QoS policer scale enhancement using statistics resource reassignment 54

Guidelines for configuring ingress QoS policer scale enhancement 55

Limitations for configuring ingress QoS policer scale enhancement 55

Configure ingress QoS policer scale enhancement using statistics resource reassignment 56

Traffic Shaping 58

Configure VOQ-Level Traffic Shaping 58

Burst Size for Port-Level Shaper 61

Burst Size for Port-Level Shaper: Why 62

Burst Size for Port-Level Shaper: Guidelines and Limitations 62

Configure Burst Size for Port-Level Shaper 63

Traffic Policing 67

Committed Bursts and Excess Bursts 68

Single-Rate Policer 68

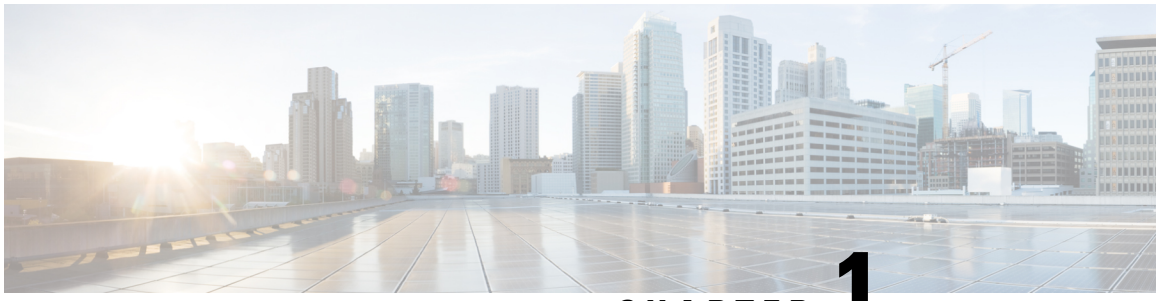
Configure Traffic Policing (Single-Rate Two-Color) 69

Configure Traffic Policing (Single-Rate Three-Color) 71
Two-Rate Policer 72
Configure Traffic Policing (Two-Rate Three-Color) 73
Packets-Per-Second-Based Policer 75
References for Modular QoS Congestion Management 78
Committed Bursts 78
Excess Bursts 79
Two-Rate Policer Details 79

CHAPTER 6 **Configuring Modular QoS on Link Bundles** 81
QoS on Link Bundles 81
Load Balancing 81
Configure QoS on Link Bundles 82

CHAPTER 7 **Configuring Hierarchical Modular QoS** 87
Overview of Hierarchical Modular QoS 87
Restrictions, Limitations, and Requirements for Configuring H-QoS 87
Configuring Hierarchical Queuing 89

CHAPTER 8 **QoS for Bridge-Group Virtual Interfaces** 95
Information on Qos on BVI 95
Restrictions on BVI 95
Classification and Marking 96
Configuring QoS on BVI 97
Verifying QoS on BVI 98



CHAPTER 1

YANG Data Models for QoS Features

This chapter provides information about the YANG data models for QoS features.

- Using YANG Data Models, on page 1

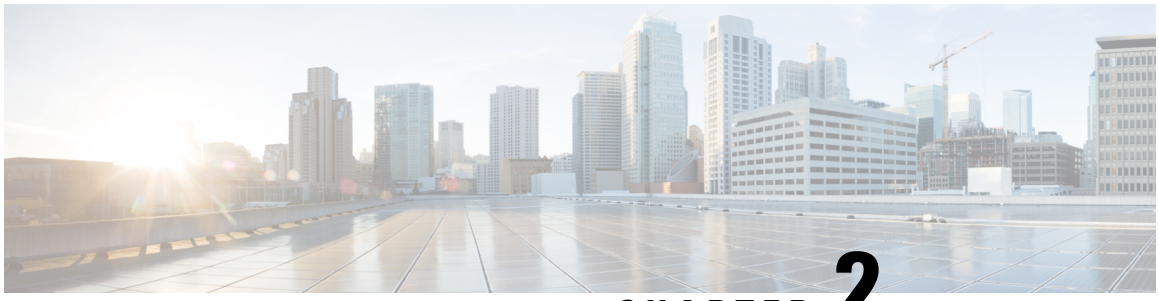
Using YANG Data Models

Cisco IOS XR supports a programmatic way of configuring and collecting operational data of a network device using YANG data models. Although configurations using CLIs are easier and human-readable, automating the configuration using model-driven programmability results in scalability.

The data models are available in the release image, and are also published in the Github repository. Navigate to the release folder of interest to view the list of supported data models and their definitions. Each data model defines a complete and cohesive model, or augments an existing data model with additional XPaths. To view a comprehensive list of the data models supported in a release, navigate to the **Available-Content.md** file in the repository.

You can also view the data model definitions using the YANG Data Models Navigator tool. This GUI-based and easy-to-use tool helps you explore the nuances of the data model and view the dependencies between various containers in the model. You can view the list of models supported across Cisco IOS XR releases and platforms, locate a specific model, view the containers and their respective lists, leaves, and leaf lists presented visually in a tree structure. This visual tree form helps you get insights into nodes that can help you automate your network.

To get started with using the data models, see the *Programmability Configuration Guide*.



CHAPTER 2

Key Hardware Resources for QoS

QoS implementations rely on various hardware resources to support efficient packet processing and traffic management.

See the table to know more about the key hardware resources that QoS utilizes.

Table 1: Key Hardware Resources that QoS Uses

Resource	Details
Switches and Routers	Switches and routers are fundamental networking devices that play a crucial role in QoS. They are responsible for forwarding traffic based on QoS policies and handling tasks such as packet classification, queuing, and scheduling.
Network Processors	Specialized hardware components designed to handle the processing and forwarding of network traffic. They typically offer high-speed packet processing capabilities, including packet classification, traffic shaping, and queuing tasks.
Traffic Shaper Devices	Hardware devices that regulate the outgoing traffic rate to meet specified bandwidth limits. They typically incorporate algorithms and buffers to shape the traffic flow and smooth out bursts or peaks in network utilization.
Application-Specific Integrated Circuits (ASICs)	ASICs are custom-designed integrated circuits tailored for specific networking tasks. They offer hardware acceleration for essential QoS functions such as packet classification, marking, queuing, and shaping, enabling high-performance and low-latency processing.
Line Cards	Components in networking devices such as switches and routers. They contain the necessary hardware resources, including ASICs, memory, and interfaces, to handle the packet processing and traffic management functions for specific ports or interfaces.

Resource	Details
Memory	Memory resources, including volatile (for example, RAM) and non-volatile (for example, flash memory), are essential for storing and managing QoS-related data, such as traffic classification rules, queuing parameters, and packet buffers.
Interfaces	Network interfaces, such as Ethernet ports or WAN interfaces, are hardware components that connect networking devices to the network. These interfaces often have QoS-related features, such as traffic shaping capabilities or support for differentiated services code points (DSCPs), enabling efficient packet processing and traffic management.

Regardless of whether you're setting up your network and plan to implement QoS or already managing your network and using QoS, it's essential to plan for these resources so you don't run out of them while, for example, deploying topologies for high-scale requirements. An essential part of this implementation is monitoring the availability of QoS resources.

- View Packet Processing and Traffic Management Resources, on page 5

View Packet Processing and Traffic Management Resources

Table 2: Feature History Table

Feature Name	Release Information	Feature Description
View Packet Processing and Traffic Management Resources	Release 7.11.1	You can now view the utilization of some packet processing and traffic management resources, such as policer banks and connectors. Insights into their consumption and availability help you prevent or mitigate an Out of Resource (OOR) situation, thus ensuring optimal QoS operations with minimal impact on network performance. The feature introduces the following changes: **CLI:** **show controllers npu resources qos** **YANG:** • ~~Cisco-IOS-XR-fretta-bcm-dpa-qos-resources-oper~~ • Cisco-IOS-XR-5500-qos-oper • ~~Cisco-IOS-XR-fretta-bcm-dpa-qos-ratepolice-resources-oper~~ • ~~Cisco-IOS-XR-fretta-bcm-dpa-qos-egresources-oper~~ (see GitHub, YANG Data Models Navigator)

QoS utilizes hardware resources to classify, mark, and shape traffic. Given that these resources are limited, you require immediate access to information such as their consumption and availability, especially if you plan to implement high-scale setups. Without such information, you may run out of resources, and tracking down the impacted hardware resources could prove challenging.

From Release 7.11.1 onwards, you can view the utilization of some packet processing and traffic management resources by running the **show controllers npu resources qos** command. Packet processing and traffic management resources are crucial components of QoS implementations, helping ensure that network traffic is prioritized, managed, and controlled effectively. You can use this data to troubleshoot Out of Resource (OOR) situations by quickly identifying the exhausted resources or prevent OOR conditions by appropriately modifying their allocation.

Run the **show controllers npu resources qos** command to view the information listed in the table. To best use this information and plan the grouping of multiple resources for specific QoS tasks, you could initially add one policy and check how much of these resources QoS utilizes.

Table 3: Packet Processing and Traffic Management Resources and their Details

Resource	Type	More About This Resource	Important Because
Policer	Packet Processing	The number of policers that are allocated and that you can create. The total number of policer banks is inversely proportional to the class map size.	Helps you determine the ingress policy map scale based on available policer bank resources.
Egress QoS Map Information	Packet Processing	The resources used in creating traffic class and drop precedence (or discard class) maps for egress traffic.	Displays counters for L2 and L3 pools, helping you tune your policy map configurations at the ingress or egress based on the existing policy parameters and optimize the available pool usage.

Resource	Type	More About This Resource	Important Because
Connectors	Traffic Management	The number of NPU connectors or Fabric Access Processors (FAPs) used.	The connector resource dictates the number of VOQs you can create. The egress policy map scale is bound to VOQ availability and, in turn, connector availability. It also tells you, among others: • the **Reserved** number of normal and low rate connector sets for internal VOQ creation; internal VOQs are used for punting and recycling operations. • the **Per Interface** counter which indicates the number of connectors used per interface or VOQ. **Note** By default, two connector regions per core are reserved for low connectors, which support lower traffic shaping rates.
Rate Profile	Traffic Management	The rate profile pool used. Interfaces with different line rates and policies with significantly differing shaper and queue lengths use rate profiles.	Helps you tune your egress policy parameters, such as shaper value and queue length, to optimize the available rate profile hardware resources.
Egress Queue (EGQ) Profile	Traffic Management	The number of egress queue maps used. They are consumed when you create egress policies with priority marking.	Helps you tune your egress policy parameters such as egress queue map to optimize the available hardware resources .

Resource	Type	More About This Resource	Important Because
VOQ	Traffic Management	The number of Virtual Output Queues (VOQs) used at a global level from a common pool.	Tells you the number of VOQs consumed (**Used**), the number of VOQs remaining (**Free**), and the **Total** number of VOQs available.

Verification

```
Router#show controllers npu resources qos all instance all location 0/0/CPU0

========================================================================
QoS TM Connectors Information For Location: 0/0/CPU0

System information for NPU 0:
Core 0:
                               Used(   %)       Free      Total   Reserved        Per
Interface Consumption
    Normal Connector Set:        384(  0%)      59008      59392       6144
             16
   Lowrate connector Set:          0(  0%)          0          0

Core 1:
                               Used(   %)       Free      Total   Reserved        Per
Interface Consumption
    Normal Connector Set:        384(  0%)      59008      59392       6144
             16
   Lowrate connector Set:          0(  0%)          0          0

========================================================================
QoS TM Rate Profile Information:
Note: - This is a global resource and shared across all LCs

                               Used(   %)       Free      Total     Reserved      Highest
 Used
                                  3(  5%)         56         59            5
  3

========================================================================
QoS TM EGQ Profile Information For Location: 0/0/CPU0

System information for NPU 0:
Core 0:
                               Used(   %)       Free      Total
                                  8(100%)          0          8

Core 1:
                               Used(   %)       Free      Total
                                  8(100%)          0          8

========================================================================
QoS PP Policer Banks Information For Location: 0/0/CPU0

System information for NPU 0:
Core 0:
                               Used(   %)       Free      Total         Class-map size
      Policer Banks info:         0(  0%)        218        218                   32
```

```
Core 1:
                                Used(   %)       Free      Total          Class-map size
      Policer Banks info:          0(  0%)        218        218                    32

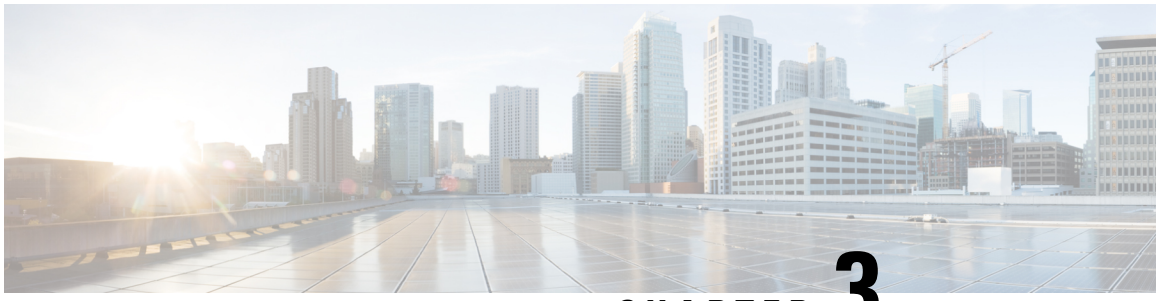
======================================================================
QoS PP Egress QoS Map Information For Location: 0/0/CPU0

System information for NPU 0:      Used(   %)       Free      Total
         L2 Profile Pool:             0(  0%)         14         14
         L3 Profile Pool:             0(  0%)          3          3

======================================================================
QoS VOQ Information For Location: 0/0/CPU0

Note:- This is a global resources shared by all NPUs

                                Used(   %)       Free      Total
               VOQ info:        175(  1%)      11984      12159
```



CHAPTER 3

Configuring Modular QoS Service Packet Classification

This chapter covers these topics:

- Packet Classification Overview, on page 11
- Traffic Class Elements, on page 12
- Traffic Policy Elements, on page 14
- In-Place Policy Modification, on page 32
- References for Modular QoS Service Packet Classification, on page 33

Packet Classification Overview

Packet classification involves categorizing a packet within a specific group (or class) and assigning it a traffic descriptor to make it accessible for QoS handling on the network. The traffic descriptor contains information about the forwarding treatment (quality of service) that the packet should receive. Using packet classification, you can partition network traffic into multiple priority levels or classes of service. The source agrees to adhere to the contracted terms and the network promises a quality of service. Traffic policers and traffic shapers use the traffic descriptor of a packet to ensure adherence to the contract.

Traffic policers and traffic shapers rely on packet classification features, such as IP precedence, to select packets (or traffic flows) traversing a router or interface for different types of QoS service. After you classify packets, you can use other QoS features to assign the appropriate traffic handling policies including congestion management, bandwidth allocation, and delay bounds for each traffic class.

The Modular Quality of Service (QoS) CLI (MQC) defines the traffic flows that must be classified, where each traffic flow is called a class of service, or class. Later, a traffic policy is created and applied to a class. All traffic not identified by defined classes fall into the category of a default class.

Guidelines and Limitations

- Ingress CoS marking is not supported on Cisco NCS 5500 Series routers; however, it is supported on Cisco NCS 5700 Series routers.

Note Ingress CoS marking is not supported on L3 interfaces because it can overwrite the Port Control Protocol (PCP) for traffic passing through the core network. To implement CoS marking, always configure it on egress L2 interfaces.

Traffic Class Elements

Default Traffic Class

Unclassified traffic (traffic that does not meet the match criteria specified in the traffic classes) is treated as belonging to the default traffic class.

If the user does not configure a default class, packets are still treated as members of the default class. However, by default, the default class has no enabled features. Therefore, packets belonging to a default class with no configured features have no QoS functionality. These packets are then placed into a first in, first out (FIFO) queue and forwarded at a rate determined by the available underlying link bandwidth. This FIFO queue is managed by a congestion avoidance technique called tail drop.

For egress classification, match on **traffic-class** (1-7) is supported. Match **traffic-class 0** cannot be configured. The class-default in the egress policy maps to **traffic-class 0**.

This example shows how to configure a traffic policy for the default class:

```
configure
 policy-map ingress_policy1
 class class-default
  police rate percent 30
 !
```

Create a Traffic Class

To create a traffic class containing match criteria, use the **class-map** command to specify the traffic class name, and then use the **match** commands in class-map configuration mode, as needed.

Guidelines

- Match criteria evaluation order: You can provide multiple values for a match type in a single line of configuration; that is, if the first value does not meet the match criteria, then the next value indicated in the match statement is considered for classification.
- Excluding values with **not** keyword: Use the **not** keyword with the **match** command to perform a match based on the values of a field that are not specified.
- Match commands are optional, but one required: All **match** commands specified in this configuration task are optional, but you must configure at least one match criterion for a class.
- **match-any** versus **match-all** behavior: If you specify **match-any**, one of the match criteria must be met for traffic entering the traffic class to be classified as part of the traffic class. This is the default. If you specify **match-all**, the traffic must match all the match criteria.

- Packet length match support: From Release 7.11.1 onwards, for the **match access-group** command, QoS classification based on the packet length field in the IPv4 and IPv6 headers is supported. Prior to this, support was not available for packet length and TTL (time to live) fields.
- Deny or empty ACLs not supported in class maps: For the **match access-group** command, when an ACL list is used within a class-map, the deny action of the ACL is ignored and the traffic is classified based on the specified ACL match parameters.

  An empty ACL (contains no rules, only remarks), when used within a class-map permits all traffic by default, and the implicit deny condition doesn't work with an empty ACL. The corresponding **class-map** matches all traffic not yet matched by the preceding traffic classes.
- Egress matching restrictions: The **traffic-class** and **discard-class** are supported only in egress direction, and these are the only match criteria supported in egress direction.
- Default class matching behavior: The egress default class implicitly matches **qos-group** 0 for marking policy and **traffic-class** 0 for queuing policy.
- Multicast and unicast egress ratio: Multicast takes a system path that is different than unicast on router, and they meet later on the egress in a multicast-to-unicast ratio of 20:80 on a per interface basis. This ratio is maintained on the same priority level as that of the traffic.
- Multicast QoS prioritization for egress: Egress QoS for multicast traffic treats traffic classes 0-5 as low-priority and traffic classes 6-7 as high priority. Currently, this is not user-configurable.
- Egress shaping not applied to high-priority multicast: Egress shaping does not take effect for multicast traffic in the high priority (HP) traffic classes. It only applies to unicast traffic.
- No egress mapping for ingress classes: If you set a traffic class at the ingress policy and do not have a matching class at egress for the corresponding traffic class value, then the traffic at ingress with this class will not be accounted for in the default class at the egress policy map.
- Only Class 0 maps to default class: Only traffic class 0 falls in the default class. A non-zero traffic class assigned on ingress but with no assigned egress queue, falls neither in the default class nor any other class.
- EDM timeout during line card reload: EDM timeouts for **show policy-map** CLI commands are expected during line card reload with scale configuration.
- Egress rate calculation variation on routers with Cisco NC57 line cards:
  - On Cisco NCS 5700 modular routers (in both compatibility and native modes), the **show policy-map interface** command may display noticeable variations for the egress rates. The ingress rates are displayed accurately.
  - This behavior is specific to routers with external TCAM and more than one NPU.
  - This behavior might not be noticeable when viewing the statistics for the first time but can become evident in subsequent checks. No action is required.

Configuration Example

You have to accomplish the following to complete the traffic class configuration:

1. Creating a class map
2. Specifying the match criteria for classifying the packet as a member of that particular class

(For a list of supported match types, see Traffic Class Elements, on page 12.)

```
Router# configure
Router(config)# class-map match-any qos-1
Router(config-cmap)# match qos-group 1
Router(config-cmap)# end-class-map
Router(config-cmap)# commit
```

Use this command to verify the class-map configuration:

```
Router#show class-map qos-1
1) ClassMap: qos-1    Type: qos
    Referenced by 2 Policymaps
```

Also see, Running Configuration, on page 17.

Also see, Verification, on page 17.

Related Topics

- Traffic Class Elements, on page 12
- Traffic Policy Elements, on page 14

Associated Commands

Traffic Policy Elements

A traffic policy contains three elements:

- Name
- Traffic class
- QoS policies

After choosing the traffic class that is used to classify traffic to the traffic policy, the user can enter the QoS features to be applied to the classified traffic.

The MQC does not necessarily require that the users associate only one traffic class to one traffic policy.

The order in which classes are configured in a policy map is important. The match rules of the classes are programmed into the TCAM in the order in which the classes are specified in a policy map. Therefore, if a packet can possibly match multiple classes, only the first matching class is returned and the corresponding policy is applied.

The router supports 32 classes per policy-map in the ingress direction and 8 classes per policy-map in the egress direction.

This table shows the supported class-actions on the router.

Supported Action Types	Direction supported on Interfaces
minimum-bandwidth	egress

Supported Action Types	Direction supported on Interfaces
bandwidth-remaining*	egress
mark	(See Packet Marking, on page 18)
police	ingress
priority	egress (level 1 to level 7)
queue-limit	egress
shape	egress
wred	egress

*Bandwidth and Bandwidth remaining configurations are not supported simultaneously within the same policy-map in H-QoS mode.

WRED supports **default** and **discard-class** options; the only values to be passed to the discard-class being 0 and 1.

Create a Traffic Policy

The purpose of a traffic policy is to configure the QoS features that should be associated with the traffic that has been classified in a user-specified traffic class or classes.

To configure a traffic class, see Create a Traffic Class, on page 12.

After you define a traffic policy with the **policy-map** command, you can attach it to one, or more interfaces to specify the traffic policy for those interfaces by using the **service-policy** command in interface configuration mode. With dual policy support, you can have two traffic policies, one marking and one queuing attached at the output. See, Attach a Traffic Policy to an Interface, on page 16.

Configuration Example

You have to accomplish the following to complete the traffic policy configuration:

1. Creating a policy map that can be attached to one or more interfaces to specify a service policy
2. Associating the traffic class with the traffic policy
3. Specifying the class-action(s) (see Traffic Policy Elements, on page 14)

```
Router# configure
Router(config)# policy-map  test-shape-1
Router(config-pmap)# class qos-1

/* Configure class-action ('shape' in this example).
Repeat as required, to specify other class-actions */
Router(config-pmap-c)# shape average percent 40
Router(config-pmap-c)# exit

/* Repeat class configuration as required, to specify other classes */
```

```
Router(config-pmap)# end-policy-map
Router(config)# commit
```

See, Running Configuration, on page 17.

See, Verification, on page 17.

Related Topics

- Traffic Policy Elements, on page 14
- Traffic Class Elements, on page 12

Associated Commands

- bandwidth
- bandwidth remaining
- class
- police
- policy-map
- priority
- queue-limit
- service-policy
- set discard-class
- set dscp
- set mpls experimental
- set precedence
- set qos-group
- shape

Attach a Traffic Policy to an Interface

After the traffic class and the traffic policy are created, you must attach the traffic policy to interface, and specify the direction in which the policy should be applied.

Note When a policy-map is applied to an interface, the transmission rate counter of each class is not accurate. This is because the transmission rate counter is calculated based on the exponential decay filter.

Configuration Example

You have to accomplish the following to attach a traffic policy to an interface:

1. Creating a traffic class and the associated rules that match packets to the class (see #unique_15)
2. Creating a traffic policy that can be attached to one or more interfaces to specify a service policy (see Create a Traffic Policy, on page 15)
3. Associating the traffic class with the traffic policy
4. Attaching the traffic policy to an interface, in the ingress or egress direction

```
Router# configure
Router(config)# interface HundredGigE 0/6/0/18
Router(config-int)# service-policy output
Router(config-int)# commit
```

Running Configuration

Verification

```
Router# show qos interface hundredGigE 0/6/0/18 output

NOTE:- Configured values are displayed within parentheses Interface HundredGigE0/6/0/18 ifh
 0x30001f8  -- output policy
NPU Id:                        3
Total number of classes:       2
Interface Bandwidth:           100000000 kbps
VOQ Base:                      11112
VOQ Stats Handle:              0x88430698
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                                =   qos-1
Egressq Queue ID                            =   11113 (LP queue)
Queue Max. BW.                              =   40329846 kbps (40 %)
Queue Min. BW.                              =   0 kbps (default)
Inverse Weight / Weight                     =   1 / (BWR not configured)
Guaranteed service rate                     =   40000000 kbps
TailDrop Threshold                          =   50069504 bytes / 10 ms (default)
WRED not configured for this class

Level1 Class                                =   class-default
Egressq Queue ID                            =   11112 (Default LP queue)
Queue Max. BW.                              =   101803495 kbps (default)
Queue Min. BW.                              =   0 kbps (default)
Inverse Weight / Weight                     =   1 / (BWR not configured)
Guaranteed service rate                     =   50000000 kbps
TailDrop Threshold                          =   62652416 bytes / 10 ms (default)
WRED not configured for this class
```

Related Topics

- Traffic Policy Elements, on page 14
- Traffic Class Elements, on page 12

Associated Commands

- service-policy

Packet Marking

The packet marking feature provides users with a means to differentiate packets based on the designated markings.

Egress packet marking

- The router supports egress packet marking. Match on **discard-class** on egress, if configured, can be used for a marking policy only.
- An egress policy is mandatory to create a user-defined header by stripping the initial 3 bits and copying the DSCP marking to the QoS group.

L2 ingress marking

- **Ingress marking limitations:** For the ingress pop operation, re-marking the customer VLAN tag (CoS, DEI) is not supported.
- **Egress traffic behavior:** The ingress 'pop VLAN' is translated to a 'push VLAN' for the egress traffic, and (CoS, DEI) marking is supported for newly pushed VLAN tags. If two VLAN tags are pushed to the packet header at the egress side, both inner and outer VLAN tags are marked. For example:

```
rewrite ingress tag pop 1 symmetric
rewrite ingress tag pop 2 symmetric
rewrite ingress tag translate 2-to-1 dot1q/dot1ad <> symmetric
```

VLAN tag details

- **Single tag operations:** When symmetrical pop 1 action is performed, the outer tag (CoS, DEI) is retained as the original frame.

IP traffic class imposition for SRv6 encapsulation

Starting Cisco IOS XR Software Release 24.2.1, you can set the IP traffic class imposition for SRv6 encapsulation using the new **set ip encapsulation class-of-service** CLI.

This feature allows you to set the DSCP traffic-class marking to 6 bits using the **hw-module profile segment-routing srv6 mode microsegment f3216 encapsulation traffic-class policy-map-extend** command. This configuration replaces the 3-bit DSCP marking set by the **hw-module profile segment-routing srv6 mode microsegment f3216 encapsulation traffic-class policy-map** command to 6 bits.

Packet Marking Guidelines and Limitations

Avoid conflicting packet marking configurations: While marking a packet, ensure you don't set the IP DSCP (using the **set dscp** command) and the MPLS experimental imposition values (using the **set mpls experimental imposition** command) for the same class map. Else, neither the DSCP remarking nor the MPLS EXP values may take effect at the ingress. This will cause, per default QoS behavior, the IP precedence values to be copied to the EXP bits on the imposed packets. Such an action could lead to unintended packets marked as high-priority by your customer being forwarded as high-priority MPLS packets in the network.

Viewing statistics and counters for egress policies: The statistics and counters for the egress marking policy cannot be viewed on the router.

QoS EXP matching for Layer 2 and Layer 3 VPN traffic:

QoS EXP matching for egress doesn't work for Layer 2 VPN and Layer 3 VPN traffic flowing from:

- Cisco NCS 5700 series line cards at ingress to Cisco NCS 5500 series line cards at the egress and
- Cisco NCS 5500 series line cards at ingress to Cisco NCS 5700 series line cards at egress.

Policy map limits for egress marking:

- For QOS EXP-Egress marking applied on a Layer 3 interface on Cisco routers, there is a limit of two unique policy maps per NPU.

  You can apply these policies to as many interfaces as your system resources allow. However, if you apply more than the permitted limit of unique policies, you may encounter unexpected failure.
- For QOS egress marking (CoS, DEI) applied on a Layer 2 interface, there is a limit of 13 unique policy-maps per NPU. If you exceed this number, you may encounter unexpected failure.
- Cisco NCS series routers do not support push or translate operations for dot1ad.

Usage guidelines for SRv6 encapsulation marking:

- The **set ip encapsulation class-of-service** command is not allowed in egress QoS policies.
- The IP encapsulation marking (**set ip encapsulation class-of-service**) and the MPLS experimental imposition marking (**set mpls experimental**) features are mutually exclusive. Both the configurations are not allowed at the same time.
- On Cisco NCS 5500 series routers, the **set ip encapsulation class-of-service** command is allowed only when the **policy-map-extend** option is enabled. On Cisco NCS 5700 series routers, the command is supported in native mode without enabling the **policy-map-extend** option.
- The **policy-map-extend** option enables 6-bit DSCP marking for both L2 and L3 modes at the same time. It is not feasible to enable the **policy-map-extend** option exclusively for L2 and L3 modes due to the shared hardware resources for L2 and L3 SRv6 traffic-class.
- To support DSCP marking in NCS 5700, the **propagate** option must be enabled in the SRv6 configurations.

Egress EXP Marking

Table 4: Feature History Table

Feature Name	Release Information	Feature Description
Egress EXP bits Marking Support on Single-Label	Release 7.11.1	With this feature support, you can now mark MPLS EXP bits on single-label scenario. You can now design the QoS for your network based on either single-label or global label. If the directly connected peer router doesn't advertise the explicit-null label, then the router encapsulates the packets with only the global label. Previously, the router was limited to mark MPLS EXP bits with a minimum of two labels. But even with a single EXP marking, you can prioritize certain types of traffic over others based on their importance.

Supported Packet Marking Operations

This table shows the supported packet marking operations.

Table 5: Supported Ingress Marking Types

Supported Mark Types	Range	Layer	Supports Unconditional Marking	Supports Conditional Marking	Description
DSCP	0-63	L3	Ingress	No	Supported flows: IP to IP, IP to MPLS, MPLS to IP.
precedence	0-7	L3	Ingress	No	Supported flows: IP to IP, IP to MPLS, MPLS to IP.
mpls experimental topmost	0-7	L3	Ingress	No	Supported flows: MPLS to MPLS.
mpls experimental imposition	0-7	L2 L3	Ingress	No	Supported flows: L2 to MPLS, IP to MPLS.
ip encapsulation	0-63	L2 L3	Ingress	No	Used for marking encapsulated SRv6 packets.
DEI	0-1	L2	Ingress	No	Supported flows: L2 to L2, L2 to MPLS.

Supported Mark Types	Range	Layer	Supports Unconditional Marking	Supports Conditional Marking	Description
CoS	0-7	L2	Ingress	No	Supported flows: L2 to L2, L2 to MPLS.
discard-class	0-2	L2 L3	Ingress	No	Used in conjunction with random-detect for discard behavior identification.
qos-group	0-7	L2 L3	Ingress	No	Marks traffic for QoS management.
traffic-class	0-7	L2 L3	Ingress	No	Used for queuing purposes.

Class-based Unconditional Packet Marking

The packet marking feature allows you to partition your network into multiple priority levels or classes of service, as follows:

- Use QoS unconditional packet marking to set the IP precedence or IP DSCP values for packets entering the network. Routers within your network can then use the newly marked IP precedence values to determine how the traffic should be treated.

  On ingress direction, after matching the traffic based on either the IP Precedence or DSCP value, you can set it to a particular discard-class. Weighted random early detection (WRED), a congestion avoidance technique, thereby uses discard-class values to determine the probability that a packet is dropped.

- Use QoS unconditional packet marking to assign MPLS packets to a QoS group. The router uses the QoS group to determine how to prioritize packets for transmission. To set the traffic class identifier on MPLS packets, use the **set traffic-class** command in policy map class configuration mode.

Note Setting the traffic class identifier does not automatically prioritize the packets for transmission. You must first configure an egress policy that uses the traffic class.

Note

- Unless otherwise indicated, the class-based unconditional packet marking for Layer 3 physical interfaces applies to bundle interfaces.

Handling QoS for Locally Originated Packets

What are Locally Originated Packets?

Locally Originated Packets (LOPs) are packets generated and transmitted by the router itself. These differ from transit packets that pass through the router. LOPs include routing protocols, management protocols, and other locally initiated control messages.

Default QoS markings for protocols

The router assigns default precedence or Differentiated Services Code Point (DSCP) values to various locally originated protocols

Protocol or application	Precedence and DSCP value	L2 PCP value
BGP, OSPF, CCM (CSM), BFD, RSVP	• Precedence: 6 • DSCP: 48	• BGP: 6 (with or without MPLS encap) • BFD: 7 (with or without MPLS encap)
Telnet, SSH	• Precedence: 2 • DSCP: 16	Not applicable
SNMP	Precedence: 0	6 (with or without MPLS encap)
SSH, SFTP	Not applicable	7 (with or without MPLS encap)

Some protocols such as BGP, RSVP, CFM, and LDP and the management protocols allow setting explicit precedence or DSCP values.

Traffic class assignment by Cisco IOS XR release

Cisco IOS XR release	Traffic class value
Before Release 7.6.1	• Locally generated control plane packets, such as IS-IS and BGP, are generated using traffic-class 6. • Locally generated BFD over Bundle (IETF) packets, which are generated on the Network Processing Unit (NPU), are generated using traffic-class 6.
From Release 7.6.1 onwards	• Locally generated control plane packets, such as IS-IS and BGP, are generated using traffic-class 7. • Locally generated BFD over Bundle (IETF) packets, which are generated on the Network Processing Unit (NPU), are generated using traffic-class 7.

Note Bidirectional Forwarding Detection (BFD) uses a DSCP value of 63 (IP-ToS 255) for single-hop sessions and IP-ToS 0 for multi-hop sessions on the NCS540 router. Due to an SDK limitation, a DSCP value of 48 is not supported.

Note Bidirectional Forwarding Detection (BFD) uses a DSCP value of 63 (IP-ToS 255) for single-hop sessions and IP-ToS 0 for multi-hop sessions on the NCS560 router. Due to an SDK limitation, a DSCP value of 48 is not supported.

Managing datapath and locally originated packets

On the router, datapath packets and injected packets aren't differentiated if both their traffic classes share the same Virtual Output Queues (VOQs). Therefore, in the case of a congested VOQ, the LOCP packets are dropped. To avoid the LOCP packets drop, Cisco recommends that you have a different traffic class for data path traffic. Alternatively, you can also specify a higher bandwidth for traffic-class 7 (if ingress traffic rate is predictable).

Optimizing traffic handling by early classification and marking

Classifying traffic helps the router to recognize traffic as a certain type and mark that traffic. By marking traffic early on its travel, you can prevent excessive reclassification later. You can mark traffic at the protocol level as shown in the following examples:

Ethernet

The following configuration shows that the outbound Control Hub packets are marked with a precedence value of 2 and EXP of 2, instead of a precedence and EXP value of 6. The SSH packets have a precedence value of 3 instead of 2.

```
ethernet cfm
  mep domain FOO service FOOBAR mep-id 1
   cos 2
ssh server dscp 24
```

BGP

```
neighbor x.x.x.x dscp
```

MPLS LDP

```
mpls ldp signalling dscp
```

Telnet

```
telnet ipv4 dscp
```

SNMP

```
snmp-server ipv4 precedence/dscp
```

Syslog

```
logging ipv4 precedence/dscp
netflow
flow exporter-map TEST dscp
```

NTP

```
ntp ipv4 precedence/dscp
ssh client dscp 56
ssh server dscp 56
```

Note By default, the router marks the Precision Time Protocol (PTP) traffic as high priority. Therefore, the need to prioritize PTP traffic in the QoS configuration is not required.

LOCPs and QoS policies

LOCPs are packets generated by the router itself, such as IS-IS, ARP, and other non-IP-based control packets. These packets are critical for network operations and are treated with high priority to ensure they are not dropped under normal circumstances.

Note By default, all LOCPs are assigned to traffic-class 7. Considering that LOCPs and LOMPs are generated by the RP, an Ingress QoS policy cannot be applied. Therefore, you must ensure that the egress QoS policy includes a class-map which matches traffic-class 7. By definition, the egress QoS policy matches all implicitly marked packets.

Key features of LOCPs

Discard priority

- LOCPs originating from the RP or LC CPU have a discard priority set in the appended Buffer Header (BHDR).
- This ensures LOCPs are not dropped internally under normal conditions.
- LOMPs do not have this discard priority and are treated as normal traffic.

Queue management

- LOCPs with discard priority are placed in an implicitly allocated high-priority queue for each physical egress interface.
- During congestion, LOCPs are not subject to Weighted Random Early Detection (WRED) or Tail-drop queue-limit operations.
- The tail-drop queue-limit must be hit before LOCP packets are dropped.

QoS policy configuration

- QoS policies can be attached to physical interfaces or sub-interfaces.
- If attached to sub-interfaces, the operator cannot attach a QoS policy to the physical interface.
- LOCPs are always sent out on the default high-priority queue of the physical interface.

Bandwidth reservation

- A minimum bandwidth of MIN (1% of interface bandwidth, 10 Mbps) is reserved for the default high-priority queue on physical interfaces without a QoS policy.

- If a QoS policy is applied, the minimum bandwidth for the high-priority queue is controlled by the configured policy.

Classification and re-marking

- LOCPs are not subject to traffic policing or re-marking.
- Non-IP LOCPs, such as IS-IS and ARP, are always sent to the high-priority queue, irrespective of the QoS policy.

Session termination prevention

- During over-subscription, LOCP drops may occur, leading to session termination.
- Proper bandwidth allocation and QoS configuration are essential to prevent such scenarios.

QoS packet marking across network layers

QoS packet marking is a networking management technique used to classify and prioritize data packets. Using this, you can enable the application of traffic handling policies such as bandwidth allocation and congestion management.

This table describes the key QoS fields in packet headers.

Table 6:

Field	Layer	Purpose
L3 DSCP	Layer 3	Classify and prioritize packets in the IP header.
MPLS EXP	Layer 2.5 (MPLS)	Carry QoS information in MPLS networks.
QoS Group	Cisco	Internal classification for flexible QoS policy application.
Traffic Class	Layer 3 (IPv6)	Define QoS policies in IPv6 headers.
L2 PCP	Layer 2	Prioritize traffic in Ethernet networks using the 802.1Q VLAN tag.

QoS packet marking across network layers is explained with and without MPLS encapsulation.

- QoS packet marking across network layers with MPLS encapsulation
- QoS packet marking across network layers without MPLS encapsulation

QoS packet marking across network layers with MPLS encapsulation

In MPLS networks, the QoS marking is done using the 3-bit MPLS EXP field in the MPLS label. This field is used to classify and prioritize traffic as it traverses the MPLS network. The EXP field is often derived from the IP precedence or DSCP value of the original packet at the ingress of the MPLS network.

This table illustrates how QoS packet marking is applied to network protocols when using MPLS encapsulation across various large density routers.

Table 7:

Router	Routing Protocol	L3 DSCP	MPLS EXP	QoS group	Traffic class	L2 PCP
Large • N540-24Q8L2DD-SYS • N540-24Q2C2DD-SYS	BFD	Cs6	6	QG6	7	6
	BFD (HW Offload)	Cs6	6	QG6	7	6
	SRPM	Cs6	6	QG6	7	6
	SRPM (NPU Offload)	Cs6	6	QG6	7	6
	ISIS	NA	6	QG6	7	6
	BGP	Cs6	6	QG6	7	6
	SSH/SFTP	Cs6	6	QG6	7	6
	NTP	Cs6	6	QG6	7	6
	TACACS	Cs6	6	QG6	7	6
	SNMP	Cs6	6	QG6	7	6
	SYSLOG	Cs6	6	QG6	7	6

This table illustrates how QoS packet marking is applied to network protocols when using MPLS encapsulation across various medium density routers.

Router	Routing Protocol	L3 DSCP	MPLS EXP	QoS group	Traffic class	L2 PCP
Medium • N540-ACC-SYS • N540X-ACC-SYS • N540-24Z8Q2C-SYS • N540X-16Z4G8Q2C-D/A • N540X-16Z8Q2C-D • N540-28Z4C-SYS-D/A • N540X-12Z16G-SYS-D/A • N540-12Z20G-SYS-D/A	BFD	Cs6	6	QG7	7	7
	BFD (HW Offload)	Cs6	6	QG7	7	7
	SRPM	Cs7	6	QG7	7	7
	SRPM (NPU Offload)	Cs7	6	QG7	7	7
	ISIS	NA	6	QG7	7	7
	BGP	Cs6	6	QG6	7	6
	SSH/SFTP	Cs6	6	QG6	7	6
	NTP	Cs6	6	QG6	7	6
	TACACS	Cs6	6	QG6	7	6
	SNMP	Cs6	6	QG6	7	6
	SYSLOG	Cs6	6	QG6	7	6
	ICMP	Cs7	6	QG7	7	6

This table illustrates how QoS packet marking is applied to network protocols when using MPLS encapsulation across various small density routers.

Router	Routing Protocol	L3 DSCP	MPLS EXP	QoS group	Traffic class	L2 PCP
Small • N540X-4Z14G2Q-D/A • N540X-8Z16G-SYS-D/A • N540X-6Z18G-SYS-D/A • N540-6Z18G-SYS-D/A • N540-6Z14S-SYS-D	BFD	Cs6	6	QG7	7	7
	BFD (HW Offload)	Cs6	6	QG7	7	7
	SRPM	Cs7	6	QG7	7	7
	SRPM (NPU Offload)	Cs7	6	QG7	7	7
	ISIS	NA	6	QG7	7	7
	BGP	Cs6	6	QG6	7	6
	SSH/SFTP	Cs6	6	QG6	7	7
	NTP	Cs6	6	QG6	7	6
	TACACS	Cs6	6	QG6	7	6
	SNMP	Cs6	6	QG6	7	6
	SYSLOG	Cs6	6	QG6	7	6
	ICMP	Cs6	6	QG6	7	6

QoS packet marking across network layers without MPLS encapsulation

The layerwise QoS packet marking refers to the classification and marking of packets at different layers to prioritize traffic based on QoS policies.

This table shows how QoS packet marking is applied without MPLS encapsulation across various large density routers.

Table 8:

Router	Routing Protocol	L3 DSCP	QoS group	Traffic class	L2 PCP
Large • N540-24Q8L2DD-SYS • N540-24Q2C2DD-SYS	BFD	Cs6	QG6	7	6
	BFD (HW Offload)	Cs6	QG6	7	6
	SRPM	Cs6	QG6	7	6
	SRPM (NPU Offload)	Cs6	QG6	7	6
	ISIS	NA	QG6	7	6
	BGP	Cs6	QG6	7	6
	SSH/SFTP	Cs6	QG6	7	6
	NTP	Cs6	QG6	7	6
	TACACS	Cs6	QG6	7	6
	SNMP	Cs6	QG6	7	6
	SYSLOG	Cs6	QG6	7	6

This table shows how QoS packet marking is applied without MPLS encapsulation across various medium density routers.

Router	Routing Protocol	L3 DSCP	QoS group	Traffic class	L2 PCP
Medium • N540-ACC-SYS • N540X-ACC-SYS • N540-24Z8Q2C-SYS • N540X-16Z4G8Q2C-D/A • N540X-16Z8Q2C-D • N540-28Z4C-SYS-D/A • N540X-12Z16G-SYS-D/A • N540-12Z20G-SYS-D/A	BFD	Cs6	QG7	7	7
	BFD (HW Offload)	Cs6	QG7	7	7
	SRPM	Cs7	QG7	7	7
	SRPM (NPU Offload)	Cs7	QG7	7	7
	ISIS	NA	QG7	7	7
	BGP	Cs6	QG6	7	6
	SSH/SFTP	Cs6	QG6	7	6
	NTP	Cs6	QG6	7	6
	TACACS	Cs6	QG6	7	6
	SNMP	Cs6	QG6	7	6
	SYSLOG	Cs6	QG6	7	6
	ICMP	Cs6	QG6	7	6

This table shows how QoS packet marking is applied without MPLS encapsulation across various small density routers.

Router	Routing Protocol	L3 DSCP	QoS group	Traffic class	L2 PCP
Small • N540X-4Z14G2Q-D/A • N540X-8Z16G-SYS-D/A • N540X-6Z18G-SYS-D/A • N540-6Z18G-SYS-D/A • N540-6Z14S-SYS-D	BFD	Cs6	QG7	7	7
	BFD (HW Offload)	Cs6	QG7	7	7
	SRPM	Cs7	QG7	7	7
	SRPM (NPU Offload)	Cs7	QG7	7	7
	ISIS	NA	QG7	7	7
	BGP	Cs6	QG6	7	6
	SSH/SFTP	Cs6	QG6	7	7
	NTP	Cs6	QG6	7	6
	TACACS	Cs6	QG6	7	6
	SNMP	Cs6	QG6	7	6
	SYSLOG	Cs6	QG6	7	6
	ICMP	Cs6	QG6	7	6

Protecting Locally Originated BFD Packets

For releases before Release 7.6.1, BFD packets are injected into traffic-class 6, with drop priority 0 (equivalent of discard-class 0). If transit traffic is also classified into traffic-class 7 and the associated egress queue is congested, BFD packets may be dropped.

From Release 7.6.1, BFD packets are injected into traffic-class 7, with drop priority 0 (equivalent of discard-class 0). If transit traffic is also classified into traffic-class 7 and the associated egress queue is congested, BFD packets may be dropped.

The recommendation is to configure transit traffic-class 7 in the ingress QoS policy with discard-class 1 or 2. You must then configure WRED in the egress QoS policy that drops these packets before dropping discard-class 0.



Note The default queue length is 16 ms.

Example

```
class-map match-any NC
 match traffic-class 6 (for releases before Release 7.6.1)
 match traffic-class 7 (from Release 7.6.1)
!
policy-map CORE-OUT
```

```
 class NC
  random-detect discard-class 1 7 ms 8 ms
```

Hardware Programming

```
RP/0/RP0/CPU0:R26-Spine2-5508#sh qos interface hu0/0/0/35 output
NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/0/0/35 ifh 0x1408  -- output policy
NPU Id:                        5
Total number of classes:       2
Interface Bandwidth:           100000000 kbps
Policy Name:                 7  wred-out
SPI Id:                        0x0
VOQ Base:                      1592
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                                =   NC
Egressq Queue ID                            =   1598 (LP queue)
Queue Max. BW.                              =   20480000 kbps (20 %)
Queue Min. BW.                              =   0 kbps (default)
Inverse Weight / Weight                     =   1 / (BWR not configured)
Guaranteed service rate                     =   20000000 kbps
Peak burst                                  =   33600 bytes (default)
TailDrop Threshold                          =   40108032 bytes / 16 ms (default)

WRED profile for Discard_Class 1
WRED Min. Threshold                         =   17563648 bytes (7 ms)
WRED Max. Threshold                         =   20054016 bytes (8 ms)

Default RED profile
WRED Min. Threshold                         =   0 bytes (0 ms)
WRED Max. Threshold                         =   0 bytes (0 ms)

WRED ECN                                    =   Disabled
```

Bundle Traffic Policies

A policy can be bound to bundles. When a policy is bound to a bundle, the same policy is programmed on every bundle member (port). For example, if there is a policer or shaper rate, the same rate is configured on every port. Traffic is scheduled to bundle members based on the load balancing algorithm.

Both ingress and egress traffic is supported. Percentage-based policies are supported.

For details, see Configure QoS on Link Bundles, on page 82.

In-Place Policy Modification

The In-Place policy modification feature allows you to modify a QoS policy even when the QoS policy is attached to one or more interfaces. A modified policy is subjected to the same checks that a new policy is subject to when it is bound to an interface. If the policy-modification is successful, the modified policy takes effect on all the interfaces to which the policy is attached. However, if the policy modification fails on any one of the interfaces, an automatic rollback is initiated to ensure that the pre-modification policy is in effect on all the interfaces.

You can also modify any class map used in the policy map. The changes made to the class map take effect on all the interfaces to which the policy is attached.

Note

- The QoS statistics for the policy that is attached to an interface are lost (reset to 0) when the policy is modified.
- When a QoS policy attached to an interface is modified, there might not be any policy in effect on the interfaces in which the modified policy is used for a short period of time.
- The system does not support the show policy-map statistics for marking policies.
- An in-place modification of an ACL does not reset the policy-map statistics counter.

Note

- For QOS EXP-Egress marking applied on a Layer 3 interface on Cisco routers, there is a limit of two unique policy-maps per NPU. When the maximum limit for policy-maps is reached and you try to modify a policy-map which is shared between different interfaces, you may get an error.
- For QOS egress marking (CoS, DEI) applied on a Layer 2 interface or on L3 sub-interface, there is a limit of 13 unique policy-maps per NPU. When the maximum limit for policy-maps is reached and you try to modify a policy-map which is shared between different interfaces, you may get an error.

Verification

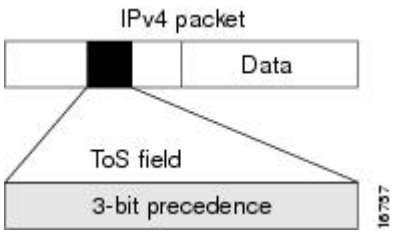
If unrecoverable errors occur during in-place policy modification, the policy is put into an inconsistent state on target interfaces. No new configuration is possible until the configuration session is unblocked. It is recommended to remove the policy from the interface, check the modified policy and then re-apply accordingly.

References for Modular QoS Service Packet Classification

Specification of the CoS for a Packet with IP Precedence

Use of IP precedence allows you to specify the CoS for a packet. You can create differentiated service by setting precedence levels on incoming traffic and using them in combination with the QoS queuing features. So that, each subsequent network element can provide service based on the determined policy. IP precedence is usually deployed as close to the edge of the network or administrative domain as possible. This allows the rest of the core or backbone to implement QoS based on precedence.

Figure 1: IPv4 Packet Type of Service Field



You can use the three precedence bits in the type-of-service (ToS) field of the IPv4 header for this purpose. Using the ToS bits, you can define up to eight classes of service. Other features configured throughout the network can then use these bits to determine how to treat the packet in regard to the ToS to grant it. These other QoS features can assign appropriate traffic-handling policies, including congestion management strategy and bandwidth allocation. For example, queuing features such as LLQ can use the IP precedence setting of the packet to prioritize traffic.

IP Precedence Bits Used to Classify Packets

Use the three IP precedence bits in the ToS field of the IP header to specify the CoS assignment for each packet. You can partition traffic into a maximum of eight classes and then use policy maps to define network policies in terms of congestion handling and bandwidth allocation for each class.

Each precedence corresponds to a name. IP precedence bit settings 6 and 7 are reserved for network control information, such as routing updates. These names are defined in RFC 791.

IP Precedence Value Settings

By default, the routers leave the IP precedence value untouched. This preserves the precedence value set in the header and allows all internal network devices to provide service based on the IP precedence setting. This policy follows the standard approach stipulating that network traffic should be sorted into various types of service at the edge of the network and that those types of service should be implemented in the core of the network. Routers in the core of the network can then use the precedence bits to determine the order of transmission, the likelihood of packet drop, and so on.

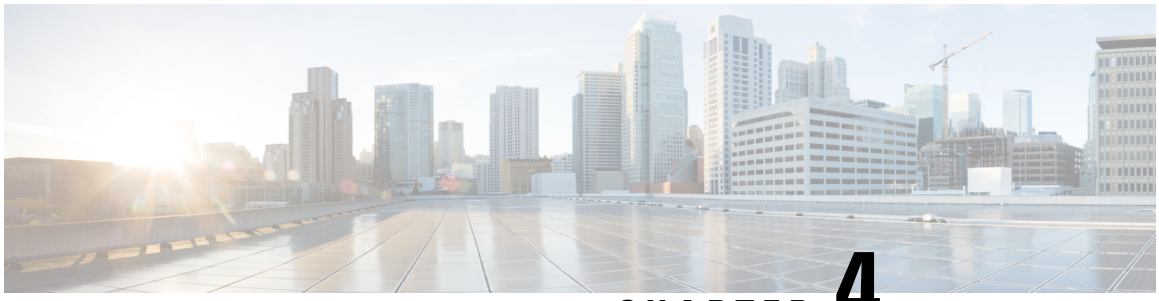
Because traffic coming into your network can have the precedence set by outside devices, we recommend that you reset the precedence for all traffic entering your network. By controlling IP precedence settings, you prohibit users that have already set the IP precedence from acquiring better service for their traffic simply by setting a high precedence for all of their packets.

The class-based unconditional packet marking and LLQ features can use the IP precedence bits.

IP Precedence Compared to IP DSCP Marking

If you need to mark packets in your network and all your devices support IP DSCP marking, use the IP DSCP marking to mark your packets because the IP DSCP markings provide more unconditional packet marking options. If marking by IP DSCP is undesirable, however, or if you are unsure if the devices in your network support IP DSCP values, use the IP precedence value to mark your packets. The IP precedence value is likely to be supported by all devices in the network.

You can set up to 8 different IP precedence markings and 64 different IP DSCP markings.



CHAPTER 4

Configuring Modular QoS Congestion Avoidance

This chapter covers the following topics:

- Modular QoS Congestion Avoidance , on page 35
- Tail Drop and the FIFO Queue, on page 36
- Random Early Detection and TCP, on page 37
- Weighted Random Early Detection, on page 40

Modular QoS Congestion Avoidance

Congestion avoidance techniques monitor traffic flow to anticipate and avoid congestion at common network bottlenecks. Avoidance techniques are implemented before congestion occurs as compared with congestion management techniques that control congestion after it has occurred.

Note For traffic requiring header decapsulation, the size of the header that is being removed is still included for the egress queuing actions. To offset this header size (required to achieve line rate for small frame sizes), configure an egress user policy with user overhead accounting on the egress interface. This policy can be a dummy policy configuration as well (allowing full traffic rate), if a policy isn't already in use or required on the egress interface.

You can enable user overhead accounting using the optional configuration of **accounting user-defined** ***<overhead size in bytes>*** while attaching the service policy on the egress interface.

Congestion avoidance is achieved through packet dropping. The router supports these QoS congestion avoidance techniques:

- Tail Drop and the FIFO Queue, on page 36
- Random Early Detection and TCP, on page 37
- Weighted Random Early Detection, on page 40

Tail Drop and the FIFO Queue

Tail drop is a congestion avoidance technique that drops packets when an output queue is full until congestion is eliminated. Tail drop treats all traffic flow equally and does not differentiate between classes of service. It manages the packets that are unclassified, placed into a first-in, first-out (FIFO) queue, and forwarded at a rate determined by the available underlying link bandwidth.

Configure Tail Drop

Packets satisfying the match criteria for a class accumulate in the queue reserved for the class until they are serviced. The **queue-limit** command is used to define the maximum threshold for a class. When the maximum threshold is reached, the enqueued packets to the class queue result in tail drop (packet drop).

Restrictions

- When configuring the **queue-limit** command, you must configure one of the following commands: **priority**, **shape average**, **bandwidth** or **bandwidth remaining**, except for the default class.

Configuration Example

You have to accomplish the following to complete the tail drop configuration:

1. Creating (or modifying) a policy map that can be attached to one or more interfaces to specify a service policy
2. Associating the traffic class with the traffic policy
3. Specifying the maximum limit the queue can hold for a class policy configured in a policy map.
4. Specifying priority to a class of traffic belonging to a policy map.
5. (Optional) Specifying the bandwidth allocated for a class belonging to a policy map or specifying how to allocate leftover bandwidth to various classes.
6. Attaching a policy map to an output interface to be used as the service policy for that interface.

```
Router# configure
Router(config)# class-map qos-1
Router(config-cmap)# match traffic-class 1
Router(config-cmap)# commit
Router(config-pmap)# exit

Router(config)# policy-map test-qlimit-1
Router(config-pmap)# class qos-1
Router(config-pmap-c)# queue-limit 100 us
Router(config-pmap-c)# priority level 7
Router(config-pmap-c)# exit
Router(config-pmap)# exit

Router(config)# interface HundredGigE 0/6/0/18
Router(config-if)# service-policy output test-qlimit-1
Router(config-if)# commit
```

Running Configuration

```
class-map qos-1
 match traffic-class 1
commit

policy-map test-qlimit-1
 class qos-1
  queue-limit 100 us
  priority level 7
 !
 class class-default
 !
 end-policy-map
!
```

Verification

```
Router# show qos int hundredGigE 0/6/0/18 output

NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/6/0/18 ifh 0x3000220  -- output policy
NPU Id:                        3
Total number of classes:       2
Interface Bandwidth:           100000000 kbps
VOQ Base:                      11176
VOQ Stats Handle:              0x88550ea0
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class (HP7)                       =   qos-1
Egressq Queue ID                         =   11177 (HP7 queue)
TailDrop Threshold                       =   1253376 bytes / 100 us (100 us)
WRED not configured for this class

Level1 Class                             =   class-default
Egressq Queue ID                         =   11176 (Default LP queue)
Queue Max. BW.                           =   101803495 kbps (default)
Queue Min. BW.                           =   0 kbps (default)
Inverse Weight / Weight                  =   1 (BWR not configured)
TailDrop Threshold                       =   1253376 bytes / 10 ms (default)
WRED not configured for this class
```

Related Topics

- Tail Drop and the FIFO Queue, on page 36

Associated Commands

- queue-limit

Random Early Detection and TCP

The Random Early Detection (RED) congestion avoidance technique takes advantage of the congestion control mechanism of TCP. By randomly dropping packets prior to periods of high congestion, RED tells the packet

source to decrease its transmission rate. Assuming the packet source is using TCP, it decreases its transmission rate until all packets reach their destination, indicating that the congestion is cleared. You can use RED as a way to cause TCP to slow transmission of packets. TCP not only pauses, but it also restarts quickly and adapts its transmission rate to the rate that the network can support.

RED distributes losses in time and maintains normally low queue depth while absorbing traffic bursts. When enabled on an interface, RED begins dropping packets when congestion occurs at a rate you select during configuration.

Configure Random Early Detection

The **random-detect** command with the **default** keyword must be used to enable random early detection (RED). The **random-detect default** command enables RED using system-defined default thresholds. These thresholds are derived dynamically based on router-specific queue characteristics and available buffer resources. Because the values are hardware-dependent, explicit minimum and maximum thresholds are not displayed or configurable when the default option is used.

Guidelines

If you configure the **random-detect default** command on any class including class-default, you must configure one of the following commands: **shape average**, **bandwidth**, and **bandwidth remaining**.

Configuration Example

You have to accomplish the following to complete the random early detection configuration:

1. Creating (or modifying) a policy map that can be attached to one or more interfaces to specify a service policy
2. Associating the traffic class with the traffic policy
3. Enabling RED with default minimum and maximum thresholds.
4. (Optional) Specifying the bandwidth allocated for a class belonging to a policy map or specifying how to allocate leftover bandwidth to various classes.
5. (Optional) Shaping traffic to the specified bit rate or a percentage of the available bandwidth.
6. Attaching a policy map to an output interface to be used as the service policy for that interface.

```
Router# configure
Router(config)# class-map qos-1
Router(config-cmap)# match traffic-class 1
Router(config-cmap)# commit
Router(config-pmap)# exit

Router# configure
Router(config)# policy-map test-wred-2
Router(config-pmap)# class qos-1
Router(config-pmap-c)# random-detect default
Router(config-pmap-c)# shape average percent 10
Router(config-pmap-c)# end-policy-map
Router(config)# commit
Router(config)# interface HundredGigE 0/6/0/18
Router(config-if)# service-policy output test-wred-2
```

```
Router(config-if)# commit
```

Running Configuration

```
class-map qos-1
 match traffic-class 1
commit

policy-map test-wred-2
 class qos-1
  random-detect default
  shape average percent 10
 !
 class class-default
 !
 end-policy-map
!

interface HundredGigE 0/6/0/18
 service-policy output test-wred-2
!
```

Verification

```
Router# show qos int hundredGigE 0/6/0/18 output

NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/6/0/18 ifh 0x3000220  -- output policy
NPU Id:                        3
Total number of classes:       2
Interface Bandwidth:           100000000 kbps
VOQ Base:                      11176
VOQ Stats Handle:              0x88550ea0
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                             =   qos-1
Egressq Queue ID                         =   11177 (LP queue)
Queue Max. BW.                           =   10082461 kbps (10 %)
Queue Min. BW.                           =   0 kbps (default)
Inverse Weight / Weight                  =   1 (BWR not configured)
Guaranteed service rate                  =   10000000 kbps
TailDrop Threshold                       =   12517376 bytes / 10 ms (default)

Default RED profile
WRED Min. Threshold                      =   12517376 bytes (10 ms)
WRED Max. Threshold                      =   12517376 bytes (10 ms)

Level1 Class                             =   class-default
Egressq Queue ID                         =   11176 (Default LP queue)
Queue Max. BW.                           =   101803495 kbps (default)
Queue Min. BW.                           =   0 kbps (default)
Inverse Weight / Weight                  =   1 (BWR not configured)
Guaranteed service rate                  =   50000000 kbps
TailDrop Threshold                       =   62652416 bytes / 10 ms (default)
WRED not configured for this class
```

Related Topics

- Random Early Detection and TCP, on page 37

Associated Commands

- random-detect

Weighted Random Early Detection

The Weighted Random Early Detection (WRED) drops packets selectively based on any specified criteria, like discard-class. WRED uses this matching criteria to determine how to treat different types of traffic.

You can configure WRED using the **random-detect** command and different discard-class values. The value can be range or a list of values that are valid for that field. You can also use minimum and maximum queue thresholds to determine the dropping point. Ensure that the WRED maximum threshold value is close to the queue limit. When the maximum threshold value is reached, packets start to get dropped.

You can also configure WRED threshold values per discard class. Such an approach helps differentiate when to drop packets among different discard classes, helping prioritize packets among discard classes.

When a packet arrives, the following actions occur:

- The average queue size is calculated.
- If the average queue size is less than the minimum queue threshold, the arriving packet is queued.
- If the average queue size is between the minimum queue threshold for that type of traffic and the maximum threshold for the interface, the packet is either dropped or queued, depending on the packet drop probability for that type of traffic.
- If the average queue size is greater than the maximum threshold, the packet is dropped.

Average Queue Size for WRED

The router automatically determines the parameters to use in the WRED calculations. The average queue size is based on the previous average and current size of the queue. The formula is:

average = (old_average * (1-2 -x)) + (current_queue_size * 2 -x)

where *x* is the exponential weight factor.

For high values of *x*, the previous average becomes more important. A large factor smooths out the peaks and lows in queue length. The average queue size is unlikely to change very quickly, avoiding a drastic change in size. The WRED process is slow to start dropping packets, but it may continue dropping packets for a time after the actual queue size has fallen below the minimum threshold. The slow-moving average accommodates temporary bursts in traffic.

Note

- The exponential weight factor, *x*, is fixed and is not user configurable.
- If the value of *x* gets too high, WRED does not react to congestion. Packets are sent or dropped as if WRED were not in effect.
- If the value of *x* gets too low, WRED overreacts to temporary traffic bursts and drops traffic unnecessarily.

For low values of *x*, the average queue size closely tracks the current queue size. The resulting average may fluctuate with changes in the traffic levels. In this case, the WRED process responds quickly to long queues. Once the queue falls below the minimum threshold, the process stops dropping packets.

Configure Weighted Random Early Detection

This configuration task is similar to that used for RED except that the **random-detect** command is not configured in RED.

Restrictions

- You cannot use the **random-detect** command in a class configured with the **priority** command, because WRED cannot be configured in a class that has been set for priority queueing (PQ).
- When configuring the **random-detect** command, you must configure one of the following commands: **shape average**, **bandwidth**, and **bandwidth remaining**.
- Do not enable Explicit Congestion Notification (ECN) and Weighted Random Early Detection (WRED) at the same time. So, do not use the following commands together:

```
*/ ECN Configuration */
Router(config-pmap-c)# random-detect ecn

*/ WRED Configuration */
Router(config-pmap-c)# random-detect discard-class 0 10 ms 500 ms
```

Configuration Example

You have to accomplish the following to complete the random early detection configuration:

1. Creating (or modifying) a policy map that can be attached to one or more interfaces to specify a service policy
2. Associating the traffic class with the traffic policy
3. Enabling WRED by specifying the match criteria (discard-class).
4. (Optional) Specifying the bandwidth allocated for a class belonging to a policy map or specifying how to allocate leftover bandwidth to various classes.
5. (Optional) Shaping traffic to the specified bit rate or a percentage of the available bandwidth.
6. (Optional) Changing queue limit to fine-tune the amount of buffers available for each queue.
7. Attaching a policy map to an output interface to be used as the service policy for that interface.

```
Router# configure
Router(config)# class-map qos-1
Router(config-cmap)# match traffic-class 1
Router(config-cmap)# commit
Router(config-pmap)# exit

Router# configure
Router(config)# policy-map test-wred-1
Router(config-pmap)# class qos-1
Router(config-pmap-c)# random-detect default
Router(config-pmap-c)# random-detect discard-class 0 10 ms 500 ms
Router(config-pmap-c)# shape average percent 10
Router(config-pmap-c)# commit

Router(config)# interface HundredGigE 0/6/0/18
Router(config-if)# service-policy output policy1
Router(config-if)# commit
```

Running Configuration

```
class-map qos-1
 match traffic-class 1
commit

policy-map test-wred-1
 class qos-1
  random-detect default
  random-detect discard-class 0 10 ms 500 ms
  shape average percent 10
 !
 class class-default
 !
 end-policy-map
!

interface HundredGigE 0/6/0/18
 service-policy output test-wred-1
!
```

Verification

```
Router# show qos int hundredGigE 0/0/0/20 output

NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/0/0/20 ifh 0x38  -- output policy
NPU Id:                        0
Total number of classes:       2
Interface Bandwidth:           100000000 kbps
Policy Name:                   test-wred-1
VOQ Base:                      1184
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                             =   qos-1
Egressq Queue ID                         =   1185 (LP queue)
Queue Max. BW.                           =   10000152 kbps (10 %)
Queue Min. BW.                           =   0 kbps (default)
Inverse Weight / Weight                  =   1 / (BWR not configured)
Guaranteed service rate                  =   10000000 kbps
Peak burst                               =   36864 bytes (default)
```

```
TailDrop Threshold                       =   1250000896 bytes / 1000 ms (default)

WRED profile for Discard_Class 0
WRED Min. Threshold                      =   12499968 bytes (10 ms)
WRED Max. Threshold                      =   624999936 bytes (500 ms)

Default RED profile
WRED Min. Threshold                      =   7499776 bytes (6 ms)
WRED Max. Threshold                      =   12499968 bytes (10 ms)

WRED ECN                                 =   Disabled

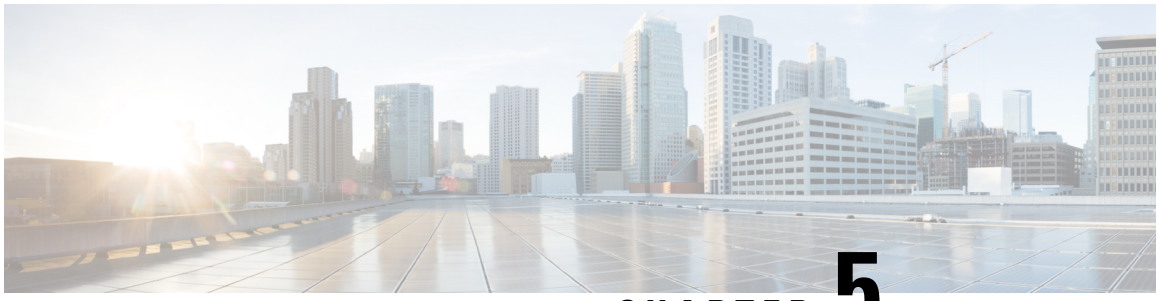
Level1 Class                             =   class-default
Egressq Queue ID                         =   1184 (Default LP queue)
Queue Max. BW.                           =   no max (default)
Queue Min. BW.                           =   0 kbps (default)
Inverse Weight / Weight                  =   1 / (BWR not configured)
Guaranteed service rate                  =   50000000 kbps
Peak burst                               =   36864 bytes (default)
TailDrop Threshold                       =   62499840 bytes / 10 ms (default)
WRED not configured for this class
```

Related Topics

- Weighted Random Early Detection, on page 40
- Configure Random Early Detection, on page 38

Associated Commands

- random-detect



CHAPTER 5

Configuring Modular QoS Congestion Management

This chapter covers the following topics:

- Congestion Management Overview, on page 45
- Best practices for minimizing packet drops during sustained maximum line-rate traffic, on page 46
- Class-based Weighted Fair Queueing, on page 46
- Low-Latency Queuing with Strict Priority Queuing, on page 49
- Overhead Accounting, on page 52
- Ingress QoS policer scale enhancement using statistics resource reassignment, on page 54
- Guidelines for configuring ingress QoS policer scale enhancement, on page 55
- Limitations for configuring ingress QoS policer scale enhancement, on page 55
- Configure ingress QoS policer scale enhancement using statistics resource reassignment, on page 56
- Traffic Shaping, on page 58
- Burst Size for Port-Level Shaper, on page 61
- Traffic Policing, on page 67
- References for Modular QoS Congestion Management, on page 78

Congestion Management Overview

Congestion management features allow you to control congestion by determining the order in which a traffic flow (or packets) is sent out an interface based on priorities assigned to packets. Congestion management entails the creation of queues, assignment of packets to those queues based on the classification of the packet, and scheduling of the packets in a queue for transmission.

The types of traffic regulation mechanisms supported are:

- Class-based Weighted Fair Queueing, on page 46
- Low-Latency Queuing with Strict Priority Queuing, on page 49
- Traffic Shaping, on page 58
- Traffic Policing, on page 67

Best practices for minimizing packet drops during sustained maximum line-rate traffic

When operating Cisco NCS 540 routers at or near 100G line rate, packet drops may occur because of congestion in the egress traffic-management layer. Although it is not possible to eliminate packet drops at sustained wire-rate, you can apply these best practices to minimize loss and maintain optimal traffic flow:

- Shape egress traffic below full line rate: Configure traffic shaping slightly below the 100G line rate, for example, at around 98G, to maintain flows within internal thresholds and avoid Virtual Output Queue (VOQ) congestion. This approach aligns with Cisco's No Drop Rate (NDR) testing guidance.
- Tune egress QoS for burst handling: Increase queue limits moderately for Best-Effort (BE) or default traffic classes, within supported limits. Apply strict queue limits to priority traffic classes to protect their performance. Enable Weighted Random Early Detection (WRED) on BE traffic to smooth packet drops and prevent bursty tail-drop events.
- Avoid sustained oversubscription at 100% utilization: Design traffic profiles with a slight margin below full utilization, especially for mixed Maximum Transmission Unit (MTU) workloads or MPLS/L2VPN flows. While some oversubscription is supported, continuous 100% load will inevitably cause packet drops.
- Align certification and benchmarking expectations: Test and certify using the NDR methodology rather than assuming strict 100% line-rate operation. This provides a more realistic and achievable performance target.

Class-based Weighted Fair Queueing

Class-based Weighted Fair Queueing (CBWFQ) allows definition of traffic classes based on customer match criteria. With CBWFQ you can define traffic classes and assign guaranteed amount of minimum bandwidth to them. CBWFQ also allows for a strict priority queue for delay-sensitive traffic.

Bandwidth Remaining

The CBWFQ algorithm derives the weight for each class from the bandwidth remaining value allocated to the class. The **bandwidth remaining** option specifies a weight for the class to the CBWFQ . After the priority-queue is serviced, the leftover bandwidth is distributed as per bandwidth remaining ratio (BWRR) or percentage. If you do not configure this command for any class, the default value of the BWRR is considered as 1 (one). In the case of **bandwidth remaining percent**, the remaining bandwidth is equally distributed among other classes, to make it 100 percentage (100%).

Restrictions

- The **bandwidth remaining** command is supported only for egress policies.

Configuring Bandwidth Remaining

Supported Platforms: Cisco NCS 540 Series Routers.

This procedure configures the minimum bandwidth and bandwidth remaining on the router

Note The **bandwidth**, **bandwidth remaining**, **shaping**, **queue-limit** and wred commands may be configured together in the same class. But, **priority** cannot be configured along with **bandwidth**, **bandwidth remaining** and wred commands.

You can configure shape average command along with priority command.

Configuration Example

You have to accomplish the following to complete the bandwidth remaining configuration:

1. Creating or modifying a policy-map that can be attached to one or more interfaces
2. Specifying the traffic class whose policy has to be created or changed
3. Allocating the leftover bandwidth for the class
4. Attaching the policy-map to an output interface

```
Router# configure
Router(config)#class-map qos-6
Router(config-cmap)#match traffic-class 4
Router(config-cmap)#exit
Router(config-cmap)#commit

Router(config)#class-map qos-5
Router(config-cmap)#match traffic-class 5
Router(config-cmap)#commit

Router(config)# policy-map test-bw-bw-rem
Router(config-pmap)# class qos-6
Router(config-pmap-c)# bandwidth percent 60
Router(config-pmap-c)# bandwidth remaining percent 60
Router(config-pmap)#class qos-5
Router(config-pmap-c)#bandwidth percent 20
Router(config-pmap-c)#bandwidth remaining percent 40
Router(config-pmap-c# exit
Router(config-pmap)# exit
Router(config)# interface HundredGigE 0/6/0/18
Router(config-if)# service-policy output test-bw-bw-rem
Router(config-if)# commit
```

Running Configuration

```
policy-map test-bw-bw-rem
 class qos-6
  bandwidth percent 60
  bandwidth remaining percent 60
 !
 class qos-5
  bandwidth percent 20
  bandwidth remaining percent 40
 !
 class class-default
```

```
 !
 end-policy-map
!

interface HundredGigE0/6/0/18
 service-policy output test-bw-bw-rem
!
```

Verification

```
Router# show qos interface HundredGigE 0/6/0/18 output

NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/6/0/18 ifh 0x3000220  -- output policy
NPU Id:                        3
Total number of classes:       3
Interface Bandwidth:           100000000 kbps
VOQ Base:                      11176
VOQ Stats Handle:              0x88550ea0
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                             =   qos-6
Egressq Queue ID                         =   11182 (LP queue)
Queue Max. BW.                           =   100824615 kbps (default)
Queue Min. BW.                           =   60494769 kbps (60 %)
Inverse Weight / Weight                  =   2 (60%)
Guaranteed service rate                  =   71881188 kbps
TailDrop Threshold                       =   90177536 bytes / 10 ms (default)
WRED not configured for this class

Level1 Class                             =   qos-5
Egressq Queue ID                         =   11181 (LP queue)
Queue Max. BW.                           =   100824615 kbps (default)
Queue Min. BW.                           =   20164923 kbps (20 %)
Inverse Weight / Weight                  =   3 (40%)
Guaranteed service rate                  =   27920792 kbps
TailDrop Threshold                       =   35127296 bytes / 10 ms (default)
WRED not configured for this class

Level1 Class                             =   class-default
Egressq Queue ID                         =   11176 (Default LP queue)
Queue Max. BW.                           =   101803495 kbps (default)
Queue Min. BW.                           =   0 kbps (default)
Inverse Weight / Weight                  =   120 (BWR not configured)
Guaranteed service rate                  =   198019 kbps
TailDrop Threshold                       =   247808 bytes / 10 ms (default)
WRED not configured for this class
```

Related Topics

- Bandwidth Remaining, on page 46

Associated Commands

- bandwidth remaining

Low-Latency Queuing with Strict Priority Queuing

The Low-Latency Queuing (LLQ) feature brings strict priority queuing (PQ) to the CBWFQ scheduling mechanism. Priority queuing (PQ) in strict priority mode ensures that one type of traffic is sent, possibly at the expense of all others. For PQ, a low-priority queue can be detrimentally affected, and, in the worst case, never allowed to send its packets if a limited amount of bandwidth is available or the transmission rate of critical traffic is high.

Configuring Low Latency Queuing with Strict Priority queuing

Configuring low latency queuing (LLQ) with strict PQ allows delay-sensitive data such as voice to be de-queued and sent before the packets in other queues are de-queued.

Support information

Supported priority levels

- Priority levels 1 to 7 are supported, with 1 being the highest and 7 the lowest.
- The default Class of Service Queue (CoSQ) 0 has the lowest priority among all levels.

Profile-based priority support

- For non-H-QoS profiles, priority levels 1 to 7 are supported.
- For H-QoS profiles, only priority levels 1 to 4 are supported.
- Regardless of the profile type, class-default always maps to CoSQ 0, which has the lowest priority.

Queue assignment for priority Level 1

- Any one of the eight egress class-maps (queues) can have priority level 1 configured.

Guidelines

Queue configuration commands

- You can configure these commands along with the **priority** command:
  - **shape average**
  - **queue-limit**
  - **random-detect**

Priority queue oversubscription

- A PQ can oversubscribe bandwidth when other queues do not fully utilize the port bandwidth.
- Oversubscription is supported only for a single priority level.

Limitations

Egress policing

- Egress policing is not supported. Hence, in strict priority queuing scenarios, lower-priority queues may not be serviced if the priority queue consumes the available bandwidth.

Multiple priority level oversubscription

- Bandwidth over-subscription with multiple priority levels is not a supported configuration.
- Starting with Cisco IOS XR Release 7.0.1, if multiple priority levels are configured for oversubscription:
  - the configuration is accepted, and a warning message is displayed
  - fair queuing is applied across all configured priority levels and,
  - higher-priority traffic is not guaranteed precedence over lower-priority traffic.

Traffic Disruption Risk

- There can be minimal traffic disruption when priority level 1 configuration is applied to any of the 8 queues.

Configuration Example

You have to accomplish the following to complete the LLQ with strict priority queuing:

1. Creating or modifying a policy-map that can be attached to one or more interfaces
2. Specifying the traffic class whose policy has to be created or changed.
3. Specifying priority to the traffic class
4. (Optional) Shaping the traffic to a specific bit rate
5. Attaching the policy-map to an output interface

```
Router# configure
Router(config)#class-map qos-1
Router(config-cmap)#match traffic-class 1
Router(config-cmap)#commit

Router(config)#class-map qos-2
Router(config-cmap)#match traffic-class 2
Router(config-cmap)#commit

Router(config)# policy-map test-priority-1
Router(config-pmap)# class qos1
Router(config-pmap-c)# priority level 7
Router(config-pmap-c)# shape average percent 2
Router(config-pmap-c)# class qos-2
Router(config-pmap-c)# priority level 6
Router(config-pmap-c)# shape average percent 1
Router(config-pmap-c)# commit
Router(config-pmap-c# exit
Router(config-pmap)# exit

Router(config)# interface HundredGigE 0/0/0/20
Router(config-if)# service-policy output test-priority-1
Router(config-if)# commit
```

Running Configuration

```
policy-map test-priority-1
 class qos-1
  priority level 7
  shape average percent 2
 !
 class qos-2
  priority level 6
  shape average percent 1
 !
 class class-default
 !
 end-policy-map
!

interface HundredGigE0/0/0/20
 service-policy output test-priority-1
!
```

Verification

```
Router#  show qos int hundredGigE 0/0/0/20 output

NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/0/0/20 ifh 0x38  -- output policy
NPU Id:                        0
Total number of classes:       3
Interface Bandwidth:           100000000 kbps
Policy Name:                   test-priority-1
VOQ Base:                      1184
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class (HP7)                          =   qos-1
Egressq Queue ID                            =   1185 (HP7 queue)
Queue Max. BW.                              =   2000000 kbps (2 %)
Guaranteed service rate                     =   2000000 kbps
Peak burst                                  =   36864 bytes (default)
TailDrop Threshold                          =   2499840 bytes / 10 ms (default)
WRED not configured for this class

Level1 Class (HP6)                          =   qos-2
Egressq Queue ID                            =   1186 (HP6 queue)
Queue Max. BW.                              =   1000000 kbps (1 %)
Guaranteed service rate                     =   1000000 kbps
Peak burst                                  =   36864 bytes (default)
TailDrop Threshold                          =   1249792 bytes / 10 ms (default)
WRED not configured for this class

Level1 Class                                =   class-default
Egressq Queue ID                            =   1184 (Default LP queue)
Queue Max. BW.                              =   no max (default)
Queue Min. BW.                              =   0 kbps (default)
Inverse Weight / Weight                     =   1 / (BWR not configured)
Guaranteed service rate                     =   97000000 kbps
Peak burst                                  =   36864 bytes (default)
TailDrop Threshold                          =   121249792 bytes / 10 ms (default)
WRED not configured for this class
```

Related Topics

- Congestion Management Overview, on page 45
- Configure VOQ-Level Traffic Shaping, on page 58
- Bandwidth Remaining, on page 46

Associated Commands

- priority

Overhead Accounting

Traffic shapers and policers use packet traffic descriptors to ensure adherence to the service level agreement in QoS. However, when traffic flows from one hop to another in a network, headers added or removed at interim hops affect the packet bytes being accounted for by QoS at each hop. When your end-user network measures the packet bytes to ensure they receive the payload as agreed, these additional header bytes cause a discrepancy.

QoS overhead accounting provides the flexibility to operators to decide which header bytes can be excluded by the traffic shaper and policer and which can be included, depending on the end user's requirements and device capabilities, to meet the committed payload in units of bytes.

For example, if the QoS commitment includes the additional header bytes, the overhead accounting feature allows your router to account for this overhead and reduces the traffic policing and shaping rates accordingly. This is also called a **positive accounting overhead**.

If however, the committed rate doesn't include the additional bytes, overhead accounting allows your router to adjust the core stream traffic such that the traffic policing and shaping rates are increased. This is also called a **negative accounting overhead**.

To summarize, QoS overhead accounting enables the router to account for packet overhead when shaping and policing traffic to a specific rate. This accounting ensures that the router runs QoS features on the actual bandwidth that the subscriber traffic consumes.

Any interface that supports QoS policies supports overhead accounting.

Note You can enable user overhead accounting using the optional configuration of **accounting user-defined** <***overhead size in bytes***> while attaching the service policy on the egress interface.

Guidelines and Restrictions

- Overhead accounting for ingress shaping is not supported.

The following restrictions apply for routers that have Cisco NC57 line cards installed and operate in native and compatibility modes.

- More than one compensation value can be programmed, provided you configure egress policy maps on different egress ports.

- You must configure a unique compensation value for a main interface and all sub-interfaces belonging to that main interface. You can't program different compensation values on different sub-interfaces sharing a common main interface.
- You can configure different compensation values on different sub-interfaces if they belong to other main interfaces.
- Compensation value programmed on egress queues (but not on VoQs) will remain active until the last egress policy map (with header compensation) is removed from main or sub-interfaces. This may impact traffic flow on main and sub-interfaces even though no compensation is set for them.

The following restrictions apply for routers that have line cards other than Cisco NC57 line cards.

- You can't program more than one compensation value per NPU or router, even if they're on different egress ports.
- You can configure the same egress compensation for different egress ports.
- NPUs can have different compensation values configured on different line cards in a modular system.
- Compensation value programmed on egress queues (but not on VoQs) will remain active until the last egress policy map (with header compensation) is removed from main or sub-interfaces. This may impact traffic flow on main and sub-interfaces even though no compensation is set for them.

Configuring for Overhead Accounting

To configure overhead accounting, you must:

1. Create a policy map and configure QoS actions for that map.
2. Configure overhead accounting and attach the map to an egress interface.

```
/* create QoS policy */
Router#configure terminal
Router(config)#policy-map policer
Router(config-pmap)#class class-default
Router(config-pmap-c)#police rate percent 10
Router(config-pmap-c-police)#commit

/* configure account overhead value while attaching the QoS policy to an egress interface
*/
Router(config)#int hundredGigE 0/0/0/2
Router(config-if)#service-policy output policer account user-defined 12
Router(config-if)#commit
Router(config-if)#root
Router(config)#end
```

Running Configuration

```
Router#sh run int hundredGigE 0/0/0/2
interface HundredGigE0/0/0/2
service-policy output policer account user-defined 12
!
```

The following example shows how to **configure a negative overhead accounting value**:

```
Router#conf
Router(config)#int hundredGigE 0/0/0/2
Router(config-if)#service-policy output policer account user-defined -12
Router(config-if)#commit
```

To **modify an overhead accounting value,** you must:

1. Remove the existing QoS policy and re-add it.
2. Configure the new overhead accounting value.

```
Router#conf
Router(config)#int hundredGigE 0/0/0/2
Router(config-if)#no service-policy input policer
Router(config-if)#service-policy output policer account user-defined -20
Router(config-if)#commit
Router#sh run int hundredGigE 0/0/0/2
interface HundredGigE0/0/0/2
service-policy output policer account user-defined -20
!
```

Positive Accounting Use Case

If QoS commitment includes Preamble, Frame Delimiter & Interframe Gap and has the following configuration:

```
service-policy output <foo> account user-defined +20
```

For QoS purposes, your router treats this packet as a packet of size = Actual Packet size + 20. Hence, the effective policing and shaping is *reduced* to match the downstream interface.

Negative Accounting Use Case

If QoS commitment to your router does not include VLAN header information, and has the following configuration:

```
service-policy output <foo> account user-defined -4
```

For QoS purposes, your router treats this packet as a packet of size = Actual Packet size – 4. Hence, the effective policing and shaping is *increased* to match the downstream interface.

Associated Commands

service-policy (overhead accounting)

Ingress QoS policer scale enhancement using statistics resource reassignment

An ingress QoS policer scale enhancement using statistics resource reassignment is a QoS policy scale feature that

- increases the maximum number of ingress QoS policers supported per system
- enables policy enforcement across a larger number of ingress L2 subinterfaces, and
- achieves higher policer scale by reallocating statistics resources from other features.

Statistics resource reallocation is the internal mechanism that enables the increased ingress policer scale. It dynamically assigns free and unassigned counter engines—which are organized as statistics banks within the NPU and used to store per-feature hardware counters—from other features to QoS Rx policer statistics. This mechanism allows the system to expand policer capacity without introducing additional hardware resources.

Table 9: Feature History Table

Feature Name	Release Information	Feature Description
Ingress QoS policer scale enhancement using statistics resource reassignment	Release 25.4.1	Introduced in this release on NCS 540 fixed port routers (select variants only*) You can now configure a maximum of 16,000 ingress QoS policers. With more QoS policers, you can apply policies on more ingress L2 subinterfaces, improving traffic control and service differentiation at scale. This is achieved by reallocating statistics resources from other features, allowing policer statistics to be assigned where scale demands are highest. Previously, the statistics reassignment command was blocked for policer statistics and could be used only by other features. *This feature is supported on N540-24Q8L2DD-SYS routers.

Guidelines for configuring ingress QoS policer scale enhancement

- Enable this feature only when high ingress policer scale is required.
- Ensure that unused statistics engines are available before reassignment.
- The QoS policer scale numbers are only applicable when the **hw-module profile mdb l2max** or **hw-module profile mdb l3max** CLI command is enabled. For more information on hardware MDB profiles, see NCS5700 MDB Profile in the *NCS 5500 System Setup and Software Installation Guide*.
- L2, L3, and tunnel interface configurations are successful even when statistics resources are not available. However, counter visibility may be affected until statistics resources are reassigned.

Limitations for configuring ingress QoS policer scale enhancement

- A statistics bank must be free before it can be reassigned.

- After statistics resource reassignment
  - multicast counters are not available, however multicast functionality is not impacted, and
  - egress ACL statistics may be impacted.

Reverting the statistics resource reassignment to the default allocation restores multicast counters to normal operation.

Configure ingress QoS policer scale enhancement using statistics resource reassignment

Procedure

Step 1 Configure the MDB L2max profile and reload the device.

Configure the hardware MDB profile to enable maximum L2 interface scale, and reload the all the nodes on the router so that the profile takes effect.

Example:

```
Router# configure
Router(config)# hw-module profile mdb l2max
Router(config)# commit
Router(config)# end
Router# reload location all
```

The device reloads with the MDB L2max profile configured.

Step 2 Verify that the MDB L2max profile is active.

After the device reloads, verify that the MDB profile for L2max is present in the running configuration.

Example:

```
Router# sh run | in hw
hw-module profile mdb l2max
```

The output confirms that the MDB L2max profile is configured on the device.

Step 3 Display statistics resource allocation and identify free counter engines.

Use this command to view the current statistics resource allocation and identify counter engines that are in the `Free` state and available for reassignment.

Example:

```
Router# show controllers npu resources stats internal instance all location 0/RP0/CPU0
System information for NPU 0:
Counter Processor Configuration Profile: Default
================================================================================
Assigned Counter Engine Resource Information
================================================================================
Feature                Eng    State   Eng   Total Counter    Core-0 Entries
                       Id             Size    Entries          InUse
================================================================================
trap-aclpolicer          0    In use     4K        2048             147
lpts-pbr                 4    In use     8K        4096              59
```

```
aclrx-mcrouterx          5    Free       8K        8192              0
acltx                    6    Free       8K        8192              0
qos rx                   8    Free       8K        4096              0
voq                      7    In use     8K        4096             51
l3rx                    10    Free      16K       16384             0
l2rx                     9    Free       8K        8192              0
l3tx                     2    Free       4K        4096              0
l3tx2                    3    In use     4K        2048             34
l2tx-mctx               11    Free      16K       16384             0
tetx-srv6loc             1    Free       4K        4096              0
POLICER1 (CHILD)        12    Reserved  16K       16384             0
POLICER2 (PARENT)       13    Reserved  16K        4096             0

Note: Free and Unassigned Counter engines can be reassigned to other features
```

You have identified one or more counter engines in the `Free` state that you can reassign to ingress QoS policer statistics.

Note
Ensure target engines are in `Free` state before reassignment.

Step 4 Reassign free counter engines to the QoS Rx feature.

This step assigns free and unassigned counter engines to the QoS Rx feature to expand ingress QoS policer statistics capacity.

Example:

```
Router# configure
Router(config)# stats-resource reassign location 0/RP0/CPU0 internal feature qosrx eng-ids
 10 11
Router(config)# commit
```

The selected counter engines are now assigned to the QoS Rx feature and available for ingress QoS policer statistics.

Step 5 Reassign a free counter engine to the l2tx-mctx feature.

Example:

```
Router# configure
Router(config)# stats-resource reassign location 0/RP0/CPU0 internal feature l2tx-mctx
eng-ids 6
Router(config)# commit
```

The selected engine is now assigned to the l2tx-mctx feature.

Step 6 Reassign a counter engine to the l3rx feature.

This step exchanges an 8K engine with a 16K engine to optimize statistics allocation.

Example:

```
Router# configure
Router(config)# stats-resource reassign location 0/RP0/CPU0 internal feature l3rx eng-ids
8
Router(config)# commit
```

The counter engine is reassigned to the L3RX feature.

Step 7 Verify the final statistics resource allocation.

Verify that the l2tx-mctx, l3rx, and QoS Rx features are using the reassigned counter engines.

Example:

```
Router# show controllers npu resources stats internal instance all location 0/RP0/CPU0
System information for NPU 0:
Counter Processor Configuration Profile: Default
================================================================================
Assigned Counter Engine Resource Information
================================================================================
Feature                Eng    State   Eng   Total Counter    Core-0 Entries
                       Id             Size    Entries          InUse
================================================================================
trap-aclpolicer          0    In use    4K        2048             147
lpts-pbr                 4    In use    8K        4096              58
aclrx-mcrouterx          5    Free      8K        8192               0
qos rx                  10    Free     16K       8192               0
qos rx                  11    Free     16K       8192               0
voq                      7    In use    8K        4096              51
l3rx                     8    Free      8K        8192               0
l2rx                     9    Free      8K        8192               0
l3tx                     2    Free      4K        4096               0
l3tx2                    3    In use    4K        2048              34
l2tx-mctx                6    Free      8K        8192               0
tetx-srv6loc             1    Free      4K        4096               0
POLICER1 (CHILD)        12    Reserved  16K      16384              0
POLICER2 (PARENT)       13    Reserved  16K       4096              0

Note: Free and Unassigned Counter engines can be reassigned to other features
```

The output confirms that the QoS Rx feature is using the reassigned statistics resources, and you can revert the reassignment to restore the default allocation if required.

Traffic Shaping

Traffic shaping allows you to control the traffic flow exiting an interface to match its transmission to the speed of the remote target interface and ensure that the traffic conforms to policies contracted for it. Traffic adhering to a particular profile can be shaped to meet downstream requirements, hence eliminating bottlenecks in topologies with data-rate mismatches.

Configure VOQ-Level Traffic Shaping

The traffic shaping performed on outgoing interfaces is done at the Layer 1 level and includes the Layer 1 header in the rate calculation.

Guidelines

- You can configure **shape average** command along with **priority** command.
- When testing under 100% line-rate traffic, you may encounter packet loss due to clock drift. This occurs when the Ethernet generator clock runs faster than the device clock on the router. To mitigate this, either reduce the test traffic load from 100% to around 99% or, if the traffic generator permits, adjust the PPM (Parts Per Million) setting to better synchronize with the device clock. These adjustments help minimize packet loss and improve test reliability.

Configuration Example

You have to accomplish the following to complete the traffic shaping configuration:

1. Creating or modifying a policy-map that can be attached to one or more interfaces
2. Specifying the traffic class whose policy has to be created or changed
3. Shaping the traffic to a specific bit rate
4. Attaching the policy-map to an output interface

```
Router# configure
Router(config)#class-map c5
Router(config-cmap)#match traffic-class 5
Router(config-cmap)#commit

Router(config)# policy-map egress_policy1
Router(config-pmap)# class c5
Router(config-pmap-c)# shape average 40 percent
Router(config-pmap-c# exit
Router(config-pmap)# exit
Router(config)# interface HundredGigE 0/1/0/0
Router(config-if)# service-policy output egress_policy1
Router(config-if)# commit
```

Running Configuration

```
class-map c5
 match traffic-class 5
commit

policy-map egress_policy1
 class c5
  shape average percent 40
 !
 class class-default
 !
 end-policy-map
!

interface HundredGigE0/6/0/18
 service-policy output egress_policy1
!
```

Verification

```
Router#  show qos interface hundredGigE 0/6/0/18 output

NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/6/0/18 ifh 0x3000220  -- output policy
NPU Id:                        3
Total number of classes:       2
Interface Bandwidth:           100000000 kbps
VOQ Base:                      11176
VOQ Stats Handle:              0x88550ea0
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
```

```
Level1 Class                              =   c5
Egressq Queue ID                          =   11177 (LP queue)
Queue Max. BW.                            =   40329846 kbps (40 %)
Queue Min. BW.                            =   0 kbps (default)
Inverse Weight / Weight                   =   1 (BWR not configured)
Guaranteed service rate                   =   40000000 kbps
TailDrop Threshold                        =   50069504 bytes / 10 ms (default)
WRED not configured for this class

Level1 Class                              =   class-default
Egressq Queue ID                          =   11176 (Default LP queue)
Queue Max. BW.                            =   101803495 kbps (default)
Queue Min. BW.                            =   0 kbps (default)
Inverse Weight / Weight                   =   1 (BWR not configured)
Guaranteed service rate                   =   50000000 kbps
TailDrop Threshold                        =   62652416 bytes / 10 ms (default)
WRED not configured for this class
```

Related Topics

- Congestion Management Overview, on page 45

Associated Commands

- shape average

Burst Size for Port-Level Shaper

Table 10: Feature History Table

Feature Name	Release Information	Feature Description
Burst Size for Port-Level Shaper	Release 7.11.1	You can now achieve a predictable and accurate burst size at the link level by configuring port-level shaper burst size, thus ensuring better adherence to traffic SLAs. Also, with the port-level shaper burst size configured in the egress policy maps, the predictability in peak burst ensures that you can configure any next-hop low-capacity device to handle these bursts. Previously, you could configure burst sizes, which impacted traffic flow only at the Virtual Output Queue (VOQ) level but didn't control packet transmission at the link level.

In relation to a port-level shaper, the burst size refers to the maximum amount of data that can be sent through the port within a short period, exceeding the configured shaping rate. When traffic is shaped at the port level, the shaping algorithm smooths the traffic flow by limiting the average rate at which data is transmitted or received. However, there are scenarios where bursts of traffic may occur, such as during periods of high network activity or when multiple packets arrive simultaneously. In such scenarios, the burst size parameter represents the maximum burst of traffic that the port-level shaper allows before it starts to drop or delay packets.

This value influences the packet transmission and peak burst to the wire, irrespective of the shaper bursts configured on different VOQs. Thus, by providing tighter control of packet flow to the wire and next-hop devices, this functionality helps you achieve accurate and predictable peak bursts.

For example, if you configure the burst size to 100 KB, the shaper allows a burst of up to 100 KB of data to be transmitted within a short period, even if it exceeds the configured shaping rate. If the burst size exceeds 100 KB, the shaper takes action to enforce the shaping rate, such as dropping or delaying packets until the traffic falls within the specified limits. As a corollary, the expected port shaping rate may not be achieved if you configure a very low burst value.

The burst size is programmed on the egress port with:

- a default value of 32 KB for all routers

Burst Size for Port-Level Shaper: Why

Because...

Previously, you could configure burst size only for VOQ-level shaping, where the shaper bandwidth and burst sizes are programmed on the Fair Queue Elements (FQEs) in your router hardware. (FQE is a mechanism that allocates bandwidth fairly among different traffic flows or queues, and the FQEs are specific to ingress queues where egress policy maps are applied.)

and..

The VOQ-level shaper burst impacts credit flow for a particular VOQ. Configuring VOQ-level burst is a best-effort technique but may not translate to the intended peak burst at the link level.

Hence...

The burst size for the port level is port or egress queue-specific, and when you configure it, you have control over packet transmission at the link level, which means you achieve an accurate peak burst in the wire. This predictability in the burst ensures that low-end devices can absorb the burst according to their capability, and there's no unexpected drop in traffic.

Burst Size for Port-Level Shaper: Guidelines and Limitations

- You must configure the port burst along with the port shaper to activate this functionality.
- The burst size is configured on the port and egress queue level.
- The default burst values are:
  - 32 KB for all routers.
- **For Hierarchical QoS (H-QoS) and egress traffic management (ETM) models:**
  - Port shaper and burst on a particular physical interface are programmed when an egress policy map with only a **class-default** configuration and a configured shaper value is applied on that interface.
  - The shaper rate on the default class is calculated as the port shaper and the burst as the port burst.
  - For NCS 5700 line cards [Mode: Compatibility; Native] and NCS 5700 fixed port routers:
    - the port shaper is not supported if no port burst is configured. This limitation does not apply to other platforms.
    - Three-level H-QoS isn't supported. This means that you can't apply two-level egress H-QoS policies on the sub-interfaces and a port shaper policy on the main interface to achieve an aggregated port level SLA in a 1+2 H-QoS or three-level H-QoS model. See Overview of Hierarchical Modular QoS, on page 87 for details about H-QoS.
- **For non-H-QoS and non-ETM models:**
  - Port level burst on a particular interface is programmed when you configure it along with port shaper.
  - You must attach a two-level egress QoS policy map to the main interface. Here, the parent shaper is taken as the port shaper.

- The shaper rate on the parent policy configured for the default class is considered as port shaper and the burst as port burst.
- For NCS 5700 line cards [Mode: Compatibility; Native] and NCS 5700 fixed port routers, the minimum port shaper is 3 Gbps by default if no port burst is configured. For other platforms, the default minimum port shaper value is 1 Gbps. If the port burst is configured, there is no limitation on the minimum port shaper value.

- The port burst value configured on the egress interface doesn't restrict the burst sizes on the VOQs or child classes.
- The parent shaper burst size doesn't restrict the child shaper burst size. Unlike the child shaper bandwidth, a child shaper burst can be greater than the parent shaper burst.
- The actual burst size programmed on the ASIC could vary from the configured value due to a hardware approximation.

Configure Burst Size for Port-Level Shaper

Scenario 1: For Cisco NCS 5700 Series Routers, NCS 5700 line cards, and NCS 5500 Series Routers, in non-H-QoS and non-ETM mode: Assign Port Burst Value

In this scenario, you specifically assign a burst value (say, 2000 bytes) along with the shaper rate (say, 2 Gbps) on a parent policy configured to the default class. These parent shaper and burst values are taken as the port shaper and burst.

1. Create a two-level egress QoS policy map by configuring the parent policy with name, for example, **port-shaper-non-hqos**.
2. Configure the **port-shaper-non-hqos** policy for the default class. There should be no other class other than the default class in the policy map.
3. To the parent policy, apply a child policy named, for example, **egress-child**, using the **service-policy** command.
4. Configure a burst value of 2000 bytes and shaper rate of 2 Gbps for the parent (**port-shaper-non-hqos**) policy.
5. Using the **class** command, specify a traffic class (say **traffic-class-1**) for the child policy (**egress-child**).
6. Using the **shape average** command, configure the shape average for **traffic-class-1** to say, 25%.
7. Attach the parent policy map to the output interface to be used as the service policy for that interface.

This configuration applies the parent shaper and burst values that are taken as the port-level shaper and burst values.

```
Router(config)# policy-map port-shaper-non-hqos
Router(config-pmap)# class class-default
Router(config-pmap-c)# service-policy egress-child
Router(config-pmap-c)# shape average 2 gbps 2000 bytes
Router(config-pmap-c)# exit
Router(config)#policy-map egress-child
Router(config-pmap)# class traffic-class-1
Router(config-pmap-c)# shape average percent 25
Router(config-pmap-c)# exit
```

```
Router(config)# interface tenGigE 0/0/0/0
Router(config-if)# service-policy output port-shaper-non-hqos
Router(config-if)# commit
```

Running Configuration

```
policy-map port-shaper-non-hqos
  class class-default
   service-policy egress-child
   shape average 2 gbps 2000 bytes
 !

 !
policy-map egress-child
 class traffic-class-1
  shape average percent 25
 !
 !

interface tenGigE 0/0/0/0
 service-policy output port-shaper-non-hqos
!
```

Verification

Run the **show qos interface** command to confirm the Peak burst value (2000 bytes) you configured for the default class.

```
Router#show qos interface tenGigE 0/0/0/0 output
NOTE:- Configured values are displayed within parentheses
Interface TenGigE0/0/0/0 ifh 0x120  -- output policy
NPU Id:                        0
Total number of classes:       3
Interface Bandwidth:           10000000 kbps
Policy Name:                   port-shaper-non-hqos
SPI Id:                        0x0
VOQ Base:                      1024
PFC enabled:                   0
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                              =   class-default
Queue Max. BW.                            =   2020026 kbps (2 gbits/sec)
Queue Min. BW.                            =   2020026 kbps (default)
Inverse Weight / Weight                   =   1 / (BWR not configured)
Peak burst                             =   2000 bytes (2000 bytes)

   Level2 Class                           =   traffic-class-1
   Egressq Queue ID                       =   1025 (LP queue)
   Queue Max. BW.                         =   505417 kbps (25 %)
   Queue Min. BW.                         =   0 kbps (default)
   Inverse Weight / Weight                =   1 / (BWR not configured)
   Guaranteed service rate                =   500000 kbps
   Peak burst                             =   33600 bytes (default)
   TailDrop Threshold                     =   626688 bytes / 10 ms (default)
   WRED not configured for this class

   Level2 Class                           =   class-default
   Egressq Queue ID                       =   1024 (Default LP queue)
   Queue Max. BW.                         =   no max (default)
   Queue Min. BW.                         =   0 kbps (default)
   Inverse Weight / Weight                =   1 / (BWR not configured)
   Guaranteed service rate                =   1000000 kbps
   Peak burst                             =   33600 bytes (default)
   TailDrop Threshold                     =   1253376 bytes / 10 ms (default)
   WRED not configured for this class
```

Scenario 2: For Cisco NCS 5700 Series Routers, NCS 5700 line cards, and NCS 5500 Series Routers, in non-H-QoS mode: Don't Assign Port Burst Value

In this scenario, you assign only the port shaper value (say, 2 Gbps) to the default class. In such cases, the burst size for port level shaper isn't activated, and the parent class has no port burst value.

1. Create a parent policy map named, for example, **port-shaper-non-hqos**.
2. Configure the **port-shaper-non-hqos** policy for the default class.
3. Apply a child policy-map named, for example, **egress-child** to the default class.
4. Shape the traffic in the default class to an average rate of 2 Gbps, limiting the egress traffic to a specific bandwidth.
5. Attach the parent policy map to the output interface to be used as the service policy for that interface.

```
Router(config)#policy-map port-shaper-non-hqos
Router(config-pmap)#class class-default
Router(config-pmap-c)#service-policy egress-child
Router(config-pmap-c)#shape average 2 gbps
Router(config-pmap-c)#exit
Router(config-pmap)#exit
Router(config)# interface tenGigE 0/0/0/0
Router(config-if)# service-policy output port-shaper-non-hqos
Router(config-if)#commit
```

Running Configuration

```
policy-map port-shaper-non-hqos
 class class-default
  service-policy egress-child
  shape average 2 gbps
 !
!
interface tenGigE 0/0/0/0
 service-policy output port-shaper-non-hqos
!
```

Verification

When you run the **show qos interface** command, you see no port burst value assigned to the parent class.

```
Router#show qos interface tenGigE 0/0/0/0 output
Sun Sep 10 20:17:16.053 UTC
NOTE:- Configured values are displayed within parentheses
Interface TenGigE0/0/0/0 ifh 0xa0  -- output policy
NPU Id:                        0
Total number of classes:       3
Interface Bandwidth:           10000000 kbps
Policy Name:                   port-shaper-non-hqos
SPI Id:                        0x0
VOQ Base:                      1024
PFC enabled:                   0
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                             =   class-default
Queue Max. BW.                           =   2100096 kbps (2 gbits/sec)
Queue Min. BW.                           =   2100096 kbps (default)
Inverse Weight / Weight                  =   1 / (BWR not configured)

   Level2 Class                             =   traffic-class-1
```

```
   Egressq Queue ID                          =   1025 (LP queue)
   Queue Max. BW.                            =   505861 kbps (25 %)
   Queue Min. BW.                            =   0 kbps (default)
   Inverse Weight / Weight                   =   1 / (BWR not configured)
   Guaranteed service rate                   =   500000 kbps
   Peak burst                                =   33600 bytes (default)
   TailDrop Threshold                        =   626688 bytes / 10 ms (default)
   WRED not configured for this class

   Level2 Class                              =   class-default
   Egressq Queue ID                          =   1024 (Default LP queue)
   Queue Max. BW.                            =   1011732 kbps (50 %)
   Queue Min. BW.                            =   0 kbps (default)
   Inverse Weight / Weight                   =   1 / (BWR not configured)
   Guaranteed service rate                   =   1000000 kbps
   Peak burst                                =   33600 bytes (default)
   TailDrop Threshold                        =   1253376 bytes / 10 ms (default)
   WRED not configured for this class
```

Scenario 3: For NCS 5700 Series Routers and NCS 5700 line cards, in H-QoS and ETM-enabled mode: Assign Port Burst Value

In this scenario, you specifically assign a burst value (say, 2000 bytes) along with the shaper rate (say, 2 Gbps) on the default class.) See Burst Size for Port-Level Shaper: Guidelines and Limitations, on page 62 for limitations that apply to this scenario.

1. Create a policy map named, for example, **port-shaper-hqos**.
2. Configure the **port-shaper-hqos** policy for the default class. The **class-default** class matches all traffic that does not match any other specific class. There should be no other class other than the default class in the policy map.
3. Configure the traffic shaping rate and the burst value for the **class-default** class.
4. Apply the policy-map named **port-shaper-hqos** as the output service policy for the interface.

```
Router (config)# policy-map port-shaper-hqos
Router (config-pmap)# class class-default
Router (config-pmap-c)# shape average 2 gbps 2000 bytes
Router (config-pmap-c)#exit
Router (config-pmap)#exit
Router(config)# interface tenGigE 0/0/0/0
Router(config-if)# service-policy output port-shaper-hqos
Router(config-if)#commit
```

Running Configuration

```
policy-map port-shaper-hqos
 class class-default
  shape average 2 gbps 2000 bytes
 !
!
interface tenGigE 0/0/0/0
 service-policy output port-shaper-hqos
!
```

Verification

Run the **show qos interface** to confirm the peak burst value (2000 bytes) you configured for the default class. Because there's no three-level H-QoS shaping support on NCS 5700 line cards [Mode: Compatibility; Native] and NCS 5700 fixed port routers, the output displays only the default class details (and not other child levels) for which you configured the shaper and burst values.

```
Router#show qos interface tenGigE 0/0/0/0 output
NOTE:- Configured values are displayed within parentheses
Interface TenGigE0/0/0/0 ifh 0xa0  -- output policy
NPU Id:                        0
Total number of classes:       1
Interface Bandwidth:           10000000 kbps
Policy Name:                   port-shaper-hqos
SPI Id:                        0x0
VOQ Base:                      1024
PFC enabled:                   0
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                              =   class-default
Egressq Queue ID                          =   1024 (Default LP queue)
Queue Max. BW.                            =   2024355 kbps (2 gbits/sec)
Queue Min. BW.                            =   0 kbps (default)
Inverse Weight / Weight                   =   60 / (BWR not configured)
Guaranteed service rate                   =   2000000 kbps
Peak burst                                =   2000 bytes (2000 bytes)
TailDrop Threshold                        =   2506752 bytes / 10 ms (default)
WRED not configured for this class
```

Traffic Policing

Traffic policing allows you to control the maximum rate of traffic sent or received on an interface and to partition a network into multiple priority levels or class of service (CoS).Traffic policing manages the maximum rate of traffic through a token bucket algorithm. The token bucket algorithm uses user-configured values to determine the maximum rate of traffic allowed on an interface at a given moment in time. The token bucket algorithm is affected by all traffic entering or leaving the interface (depending on where the traffic policy with traffic policing is configured) and is useful in managing network bandwidth in cases where several large packets are sent in the same traffic stream. By default, the configured bandwidth value takes into account the Layer 2 encapsulation that is applied to traffic leaving the interface.

Traffic policing also provides a certain amount of bandwidth management by allowing you to set the burst size (Bc) for the committed information rate (CIR). See, Committed Bursts and Excess Bursts, on page 68.

The router supports the following traffic policing mode(s):

- Single-Rate Two-Color (SR2C) in color-blind mode. See Single-Rate Policer, on page 68.
- Single-Rate Three-Color (SR3C) in color-blind mode.
- Two-Rate Three-Color (2R3C) in color-blind mode. See Two-Rate Policer, on page 72 .

Restrictions

- Traffic policing is supported only in ingress direction, and only color-blind mode is supported.
- The policing rate accuracy may vary up to +/-2% from the configured policer value.
- Ensure that you don’t configure a policer and match criteria for **discard-class** in the same class. Even though the configuration is allowed, the policer doesn’t work and allows all traffic without dropping packets.
- Policer marking is not supported.

- Policers are configured in the interface at the core level and “show qos int <>” value is displayed at the NPU level.

  For policers configured in a bundle interface where bundle members are from the same NPU but different cores (NPU cores), each member sends the traffic up to the core level policer configuration, but “show qos int <>” displays the NPU level policer output.

- Example:

  For bundle interface with two 10GE members (same NPU, but one interface from core0, one interface from core1) 2R3C policer applied on bundle interface (1G confirm rate , 1G exceed rate – total 2G policer rate) will be shown on the “show qos int <>” output):

  Interface in core0 – 500 Mbps confirm rate, 500 Mbps exceed rate

  Interface in core1 – 500 Mbps confirm rate, 500 Mbps exceed rate

  For traffic in one out of two interfaces, the policed rate will be 1Gbps. For traffic on two interfaces, policed rate will be 2Gbps.

Committed Bursts and Excess Bursts

Unlike a traffic shaper, a traffic policer does not buffer excess packets and transmit them later. Instead, the policer executes a “send or do not send” policy without buffering. Policing uses normal or committed burst (bc) values and excess burst values (be) to ensure that the router reaches the configured committed information rate (CIR). Policing decides if a packet conforms or exceeds the CIR based on the burst values you configure. Burst parameters are based on a generic buffering rule for routers, which recommends that you configure buffering to be equal to the round-trip time bit-rate to accommodate the outstanding TCP windows of all connections in times of congestion. During periods of congestion, proper configuration of the excess burst parameter enables the policer to drop packets less aggressively.

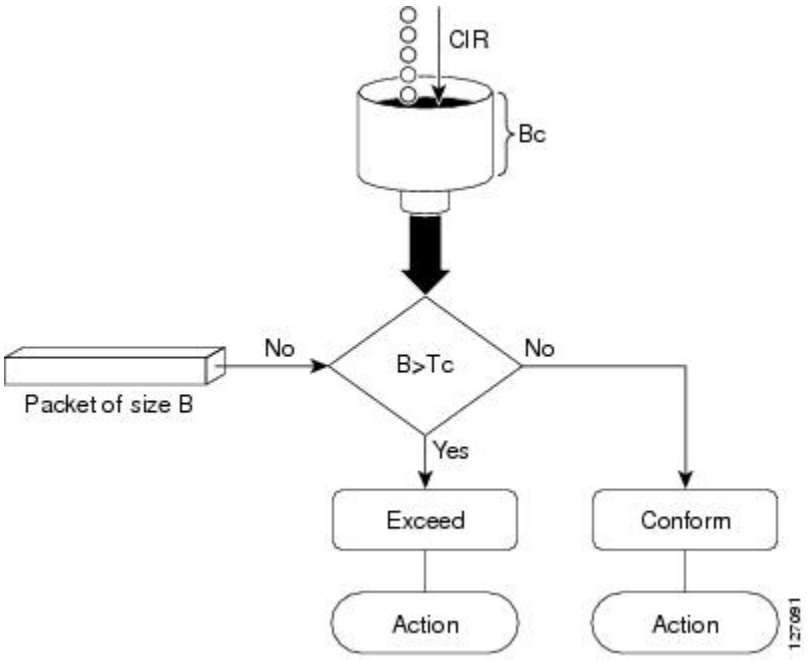
For more details, see Committed Bursts, on page 78 and Excess Bursts, on page 79.

Single-Rate Policer

Single-Rate Two-Color Policer

A single-rate two-color (SR2C) policer provides one token bucket with two actions for each packet: a conform action and an exceed action.

Figure 2: Workflow of Single-Rate Two-Color Policer



Based on the committed information rate (CIR) value, the token bucket is updated at every refresh time interval. The Tc token bucket can contain up to the Bc value, which can be a certain number of bytes or a period of time. If a packet of size B is greater than the Tc token bucket, then the packet exceeds the CIR value and a default action is performed. If a packet of size B is less than the Tc token bucket, then the packet conforms and a different default action is performed.

Single-Rate Three-Color Policer

A single-rate three-color (SR3C) policer provides one token bucket with three actions for each packet: a conform action, an exceed action and a violate action. The packet is marked based on the CIR value and the two associated burst size - committed burst size (CBS) and excess burst size (EBS). If a packet does not exceed the CBS, it is marked as conformed packet. The packet is marked as exceeded if it exceeds CBS, but not the EBS. If it exceeds the EBS as well, it is marked as violate packet.

Configure Traffic Policing (Single-Rate Two-Color)

Traffic policing is often configured on interfaces at the edge of a network to limit the rate of traffic entering or leaving the network. The default conform action for single-rate two color policer is to transmit the packet and the default exceed action is to drop the packet. Users cannot modify these default actions.

Configuration Example

You have to accomplish the following to complete the traffic policing configuration:

1. Creating or modifying a policy-map that can be attached to one or more interfaces
2. Specifying the traffic class whose policy has to be created or changed
3. (Optional) Specifying the marking action
4. Specifying the policy rate for the traffic

5. Attaching the policy-map to an input interface

```
Router# configure
Router(config)# policy-map test-police-1
Router(config-pmap)# class  ipv6-6
Router(config-pmap-c)# set dscp cs2 (optional)
Router(config-pmap-c)# set qos-group 7 (optional)
Router(config-pmap-c)# police rate percent 20 burst 10000 bytes
Router(config-pmap-c-police)# exit
Router(config-pmap-c)# exit
Router(config-pmap)# exit
Router(config)# interface HundredGigE 0/6/0/18
Router(config-if)# service-policy input test-police-1
Router(config-if)# commit
```

Running Configuration

```
class-map match-any ipv6-6
 match precedence 3
 end-class-map
!

policy-map test-police-1
 class ipv6-6
  set dscp cs2
  set qos-group 7
  police rate percent 20 burst 10000 bytes
  !
 !
 class class-default
 !
 end-policy-map
!

interface HundredGigE0/6/0/18
 service-policy input test-police-1
 service-policy output test-priority-1
!
```

Verification

```
Router# show qos interface hundredGigE 0/6/0/18 input

NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/6/0/18 ifh 0x3000220  -- input policy
NPU Id:                        3
Total number of classes:       2
Interface Bandwidth:           100000000 kbps
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                             =   ipv6-6
New dscp                                 =   16
New qos group                            =   7

Policer Bucket ID                        =   0x102a0
Policer Stats Handle                     =   0x8a8090c0
Policer committed rate                   =   19980000 kbps (20 %)
```

```
Policer conform burst                    =   9856 bytes (10000 bytes)

Level1 Class                             =   class-default

Default Policer Bucket ID                =   0x102a1
Default Policer Stats Handle             =   0x8a808e78
Policer not configured for this class
```

Related Topics

- Traffic Policing, on page 67

Associated Commands

- police rate

Configure Traffic Policing (Single-Rate Three-Color)

The default conform action and exceed actions for single-rate three-color policer are to transmit the packet and the default violate action is to drop the packet. User cannot modify these default actions.

Configuration Example

You have to accomplish the following to complete the traffic policing configuration:

1. Creating or modifying a policy-map that can be attached to one or more interfaces
2. Specifying the traffic class whose policy has to be created or changed
3. (Optional) Specifying the marking action
4. Configuring the policy rate for the traffic along with the peak-burst values
5. Attaching the policy-map to an input interface

```
Router# configure
Router(config)# policy-map test-police-1R3C
Router(config-pmap)# class ipv4-5
Router(config-pmap-c)# set qos-group 2 (optional)
Router(config-pmap-c)# police rate percent 20 burst 100000 bytes peak-burst 190000 bytes
Router(config-pmap-c-police)# exit
Router(config-pmap-c)# exit
Router(config-pmap)# exit
Router(config)# interface HundredGigE 0/6/0/18
Router(config-if)# service-policy input test-police-1R3C
Router(config-if)# commit
```

Running Configuration

```
class-map match-any ipv4-5
 match precedence 3
 end-class-map
!

policy-map test-police-1R3C
```

```
 class ipv4-5
  set qos-group 7
  police rate percent 20 burst 100000 bytes peak-burst 190000 bytes
  !
 !
 class class-default
 !
 end-policy-map
!

interface HundredGigE0/6/0/18
 service-policy input test-police-1R3C
 service-policy output test-priority-1
!
```

Verification

```
Router# show qos interface hundredGigE 0/6/0/18 input

NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/6/0/18 ifh 0x3000220  -- input policy
NPU Id:                        3
Total number of classes:       2
Interface Bandwidth:           100000000 kbps
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                              =   ipv4-5
New qos group                             =   2

Policer Bucket ID                         =   0x102a1
Policer Stats Handle                      =   0x8a8090c0
Policer committed rate                    =   19980000 kbps (20 %)
Policer conform burst                     =   99584 bytes (100000 bytes)
Policer exceed burst                      =   188672 bytes (190000 bytes)

Level1 Class                              =   class-default

Default Policer Bucket ID                 =   0x102a1
Default Policer Stats Handle              =   0x8a808e78
Policer not configured for this class
```

Related Topics

- Traffic Policing, on page 67

Associated Commands

- police rate

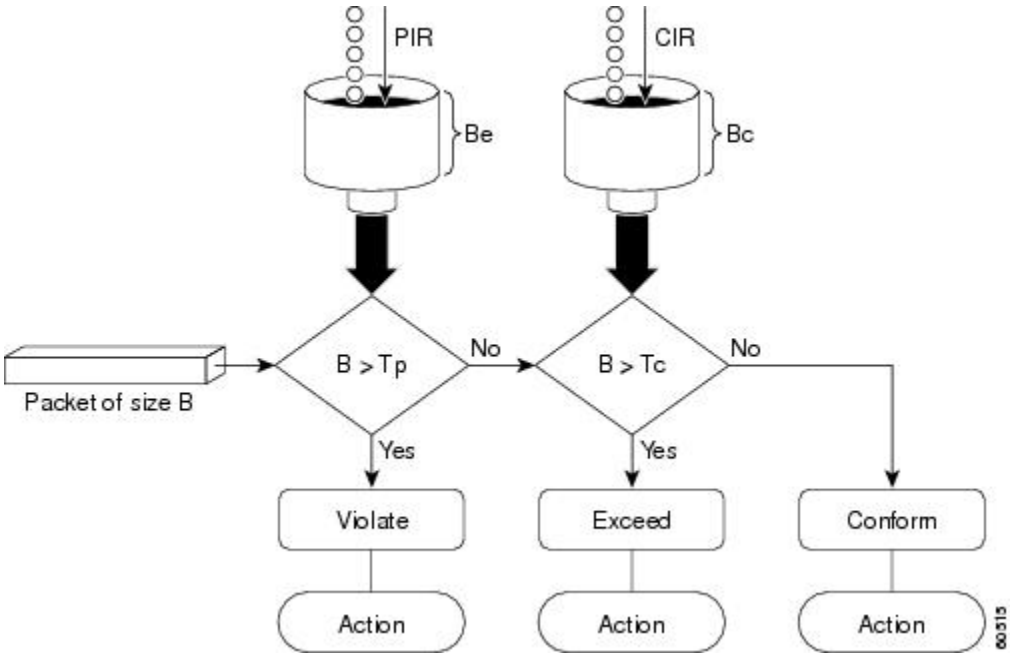
Two-Rate Policer

The two-rate policer manages the maximum rate of traffic by using two token buckets: the committed token bucket and the peak token bucket. The dual-token bucket algorithm uses user-configured values to determine the maximum rate of traffic allowed on a queue at a given moment. In this way, the two-rate policer can meter traffic at two independent rates: the committed information rate (CIR) and the peak information rate (PIR).

The dual-token bucket algorithm provides users with three actions for each packet—a conform action, an exceed action, and an optional violate action. Traffic entering a queue with the two-rate policer configured is placed into one of these categories. The actions are pre-determined for each category. The default conform and exceed actions are to transmit the packet, and the default violate action is to drop the packet.

This figure shows how the two-rate policer marks a packet and assigns a corresponding action to the packet.

Figure 3: Marking Packets and Assigning Actions—Two-Rate Policer



Also, see Two-Rate Policer Details, on page 79.

The router supports Two-Rate Three-Color (2R3C) policer.

Configure Traffic Policing (Two-Rate Three-Color)

The default conform and exceed actions for two-rate three-color (2R3C) policer are to transmit the packet and the default violate action is to drop the packet. Users cannot modify these default actions.

Configuration Example

You have to accomplish the following to complete the two-rate three-color traffic policing configuration:

1. Creating or modifying a policy-map that can be attached to one or more interfaces
2. Specifying the traffic class whose policy has to be created or changed
3. Specifying the packet marking
4. Configuring two rate traffic policing
5. Attaching the policy-map to an input interface

```
Router# configure
Router(config)# policy-map policy1
```

```
Router(config-pmap)# class ipv4-7
Router(config-pmap-c)# set qos-group 4
Router(config-pmap-c)# police rate percent 20 burst 100000 bytes peak-rate percent 50
peak-burst 200000 bytes
Router(config-pmap-c-police)# exit
Router(config-pmap-c)# exit
Router(config-pmap)# exit
Router(config)# interface HundredGigE 0/6/0/18
Router(config-if)# service-policy input policy1
Router(config-if)# commit
```

Running Configuration

```
policy-map policy1
 class ipv4-7
  set qos-group 4
  police rate percent 20 burst 100000 bytes peak-rate percent 50 peak-burst 200000 bytes
  !
 !

interface HundredGigE 0/6/0/18
 service-policy input policy1
!
```

Verification

```
Router# show policy-map interface HundredGigE 0/6/0/18

NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/6/0/18 ifh 0x3000220  -- input policy
NPU Id:                        3
Total number of classes:       8
Interface Bandwidth:           100000000 kbps
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                             =   ipv4-4
- - -
- - -
Level1 Class                             =   ipv4-7
New qos group                            =   4

Policer Bucket ID                        =   0x102a3
Policer Stats Handle                     =   0x8a8089e8
Policer committed rate                   =   19980000 kbps (20 %)
Policer peak rate                        =   49860000 kbps (50 %)
Policer conform burst                    =   99584 bytes (100000 bytes)
Policer exceed burst                     =   199168 bytes (200000 bytes)

Level1 Class                             =   class-default

Policer Bucket ID                        =   0x102a7
Policer Stats Handle                     =   0x8a7c8510
Policer committed rate                   =   29880000 kbps (30 %)
Policer conform burst                    =   4194304 bytes (default)
```

Important Notes

- A policer is programmed per NPU core on a bundle interface. So, all members on a bundle interface from the same core share the policer.

Related Topics

- Two-Rate Policer, on page 72

Associated Commands

- police rate

Packets-Per-Second-Based Policer

Table 11: Feature History Table

Feature Name	Release Information	Feature Description
Packets-Per-Second-Based Policer	Release 7.4.1	Prior to this functionality, when configuring policers, the only available option for policer rates was bit-rate measured in units of bits per second (bps). With this release, you can configure policer rates in units of packets per second (pps) as well. pps-based policer is critical in fending off malicious attacks—when attackers target your specific resources with a vast amount of traffic that contain higher number of packets, but move at a slower packet rate. Protection from such attacks is possible because pps-based policers ensure that regardless of the packet size and rate, the policer only accepts a fixed number of packets per second. This functionality modifies the **police rate** command.

- **Policer rates so far**—You used the **police rate** command to configure policers based on two parameters:
  - bit-rates (default unit: bits per second or bps)
  - Burst size (default unit: bytes)
- **packets-per-second (pps)-based policer**—With this additional functionality, you can use the **police rate** command to configure policers in units of packets per second (pps). The pps configuration option is available as **police rate <*pps-value*> pps**. When you configure the pps option, ensure that you configure burst size in packets. (See **Restrictions and guidelines**.) Thus, the parameters for pps-based policer are:

- packets per second (pps)
- burst size (packets)

- **Why pps-based-policer**—Networks face newer types of attacks, and these days malicious operators don't necessarily employ aggressive tactics that involve overwhelming your bandwidth with large amount of traffic to cause distributed denial of service (DDoS). Now, some attackers go the 'softer' route, where they send smaller packet sizes at slower traffic rates. During such malicious network activity, a bandwidth-based policer can still aggregate up to many packets to be processed if the packet size is small. Attackers tend to use this behavior to bypass bandwidth-based policers to exploit vulnerabilities or try to hit performance limitations by increasing the packet rates.

  Packets-per-second-based policers ensure that regardless of the packet size and traffic rate, the policer only accepts a fixed number of packets per second.

  pps-based-policer support cheat-sheet—Here's a quick look at some key support areas and their details for pps-based policer.

Support	Details
Classification and marking support	Same as that for bps-based-policer
Units	Equivalent kbps values display for QoS programming and statistics.
H-QoS	Support for parent and child policers
Bursts	Support for confirm burst (bc) and exceed burst (be) values in units of packets. The default value is in multiple of 128 bytes equivalent to 10 milliseconds.
Minimum pps value	For better granularity, recommended minimum value is 100 pps.

- **Restriction and guidelines**
  - This functionality is **applicable only for ingress.**
  - When using a pps-based policer, **ensure that you configure the burst-size value in number of packets as well.** This is because a policer burst rate determines whether a specific number of packets out of contract would be subject to the next action (that is, exceed or violate).
  - **Within a QoS policy, configure the parent and child policies policers to either bps or pps.** Else, the configuration displays an error when you try attaching the policy to an interface.
  - **For single-level policy maps:** under the same policy map, you can configure one class map with bps-based policer and the other class map with a pps-based policer.
  - **For two-level hierarchical policy maps:**
    - The parent and child-level policy maps must use the same unit-based policer. That is, both must have either pps-based or bps-based policers.
    - If you configure the child-level policy map with pps-based policer, ensure that the parent policy-map class default has a pps-based policer.

- **Configure pps-based policer**—To configure pps-based policer, you must:
  1. Configure a class map.
  2. Create a service policy for the map and configure the pps values.
  3. Attach the service policy to an interface.

```
/*Configure a class map*/

Router(config)#class-map prec1
Router(config-cmap)#match precedence 1
Router(config-cmap)# exit
Router(config)# commit

/*Create a service policy map*/

Router(config)# policy-map policy1
Router(config-pmap)# class prec1
Router(config-pmap-c)#police rate 1000 pps burst 300 packets
Router(config-pmap-c-police)#exit
Router(config-pmap-c)#exit
Router(config-pmap)#exit
Router(config)# commit

/*Attach the service policy to an interface*/
Router#int hundredGigE 0/7/0/2
Router(config-if)service-policy input policy1
Router(config-if)#exit
Router(config)#commit
```

Running Configuration

```
class-map match-any prec1
match precedence 1
 end-class-map
!
policy-map policy1
class prec1
  police rate 1000 pps burst 300 packets
  !
 !
 class class-default
!
 end-policy-map
!
```

Verification

```
Router#show qos int hundredGigE 0/7/0/2 input
NOTE:- Configured values are displayed within parentheses
Interface HundredGigE0/7/0/2 ifh 0xe000088  -- input policy
NPU Id:                        0
Total number of classes:       2
Interface Bandwidth:           100000000 kbps
Policy Name:                   policy1
SPI Id:                        0x0
Accounting Type:               Layer2 (Include Layer 2 encapsulation and above)
------------------------------------------------------------------------------
Level1 Class                              =   prec1

Policer Bucket ID                         =   0x9
```

```
Policer Stats Handle                    =   0x0
Policer committed rate                  =   998 kbps (1000 packets/sec)
Policer conform burst                   =   37632 bytes (300 packets)

Level1 Class                            =   class-default

Default Policer Bucket ID               =   0x8
Default Policer Stats Handle            =   0x0
Policer not configured for this class
```

Associated Commands

police rate

References for Modular QoS Congestion Management

Committed Bursts

The committed burst (bc) parameter of the police command implements the first, conforming (green) token bucket that the router uses to meter traffic. The bc parameter sets the size of this token bucket. Initially, the token bucket is full and the token count is equal to the committed burst size (CBS). Thereafter, the meter updates the token counts the number of times per second indicated by the committed information rate (CIR).

The following describes how the meter uses the conforming token bucket to send packets:

- If sufficient tokens are in the conforming token bucket when a packet arrives, the meter marks the packet green and decrements the conforming token count by the number of bytes of the packet.
- If there are insufficient tokens available in the conforming token bucket, the meter allows the traffic flow to borrow the tokens needed to send the packet. The meter checks the exceeding token bucket for the number of bytes of the packet. If the exceeding token bucket has a sufficient number of tokens available, the meter marks the packet.

  Green and decrements the conforming token count down to the minimum value of 0.

  Yellow, borrows the remaining tokens needed from the exceeding token bucket, and decrements the exceeding token count by the number of tokens borrowed down to the minimum value of 0.
- If an insufficient number of tokens is available, the meter marks the packet red and does not decrement either of the conforming or exceeding token counts.

Note When the meter marks a packet with a specific color, there must be a sufficient number of tokens of that color to accommodate the entire packet. Therefore, the volume of green packets is never smaller than the committed information rate (CIR) and committed burst size (CBS). Tokens of a given color are always used on packets of that color.

Excess Bursts

The excess burst (be) parameter of the police command implements the second, exceeding (yellow) token bucket that the router uses to meter traffic. The exceeding token bucket is initially full and the token count is equal to the excess burst size (EBS). Thereafter, the meter updates the token counts the number of times per second indicated by the committed information rate (CIR).

The following describes how the meter uses the exceeding token bucket to send packets:

- When the first token bucket (the conforming bucket) meets the committed burst size (CBS), the meter allows the traffic flow to borrow the tokens needed from the exceeding token bucket. The meter marks the packet yellow and then decrements the exceeding token bucket by the number of bytes of the packet.
- If the exceeding token bucket does not have the required tokens to borrow, the meter marks the packet red and does not decrement the conforming or the exceeding token bucket. Instead, the meter performs the exceed-action configured in the police command (for example, the policer drops the packets).

Two-Rate Policer Details

The committed token bucket can hold bytes up to the size of the committed burst (bc) before overflowing. This token bucket holds the tokens that determine whether a packet conforms to or exceeds the CIR as the following describes:

- A traffic stream is conforming when the average number of bytes over time does not cause the committed token bucket to overflow. When this occurs, the token bucket algorithm marks the traffic stream green.
- A traffic stream is exceeding when it causes the committed token bucket to overflow into the peak token bucket. When this occurs, the token bucket algorithm marks the traffic stream yellow. The peak token bucket is filled as long as the traffic exceeds the police rate.

The peak token bucket can hold bytes up to the size of the peak burst (be) before overflowing. This token bucket holds the tokens that determine whether a packet violates the PIR. A traffic stream is violating when it causes the peak token bucket to overflow. When this occurs, the token bucket algorithm marks the traffic stream red.

For example, if a data stream with a rate of 250 kbps arrives at the two-rate policer, and the CIR is 100 kbps and the PIR is 200 kbps, the policer marks the packet in the following way:

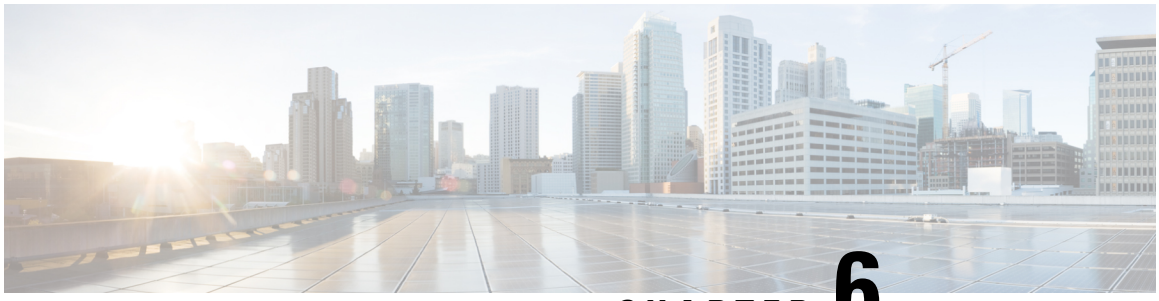
- 100 kbps conforms to the rate
- 100 kbps exceeds the rate
- 50 kbps violates the rate

The router updates the tokens for both the committed and peak token buckets in the following way:

- The router updates the committed token bucket at the CIR value each time a packet arrives at the interface. The committed token bucket can contain up to the committed burst (bc) value.
- The router updates the peak token bucket at the PIR value each time a packet arrives at the interface. The peak token bucket can contain up to the peak burst (be) value.
- When an arriving packet conforms to the CIR, the router takes the conform action on the packet and decrements both the committed and peak token buckets by the number of bytes of the packet.

- When an arriving packet exceeds the CIR, the router takes the exceed action on the packet, decrements the committed token bucket by the number of bytes of the packet, and decrements the peak token bucket by the number of overflow bytes of the packet.
- When an arriving packet exceeds the PIR, the router takes the violate action on the packet, but does not decrement the peak token bucket.

See Two-Rate Policer, on page 72.



CHAPTER 6

Configuring Modular QoS on Link Bundles

This chapter covers the following topics:

- QoS on Link Bundles, on page 81

QoS on Link Bundles

A bundle is a group of one or more ports that are aggregated together and treated as a single link. The router supports Ethernet interfaces and VLAN interfaces (bundle sub-interfaces) bundles. All QoS features currently supported on physical interfaces, are also supported on all link bundle interfaces. Applying QoS on bundle members is not supported.

Restrictions for Link Bundles

- Only Ethernet link bundling is supported.
- A bundle interface can only contain physical interface.
- All links within a single bundle must be configured either to run 802.3ad (LACP) or Etherchannel (non-LACP). Mixed links within a single bundle are not supported.
- MAC accounting is not supported on Ethernet link bundles.
- Maximum number of links supported in each link bundle is 64.
- The maximum number of link bundles supported is 128.
- On NCS-57C3-MOD-SYS routers, configuring an egress QoS service policy on a Bundle-Ether interface with a shaper burst size smaller than the interface MTU can cause the bundle interface to go down and stop forwarding traffic.

  In this condition, router reloads, line card reloads, and removal or reapplication of the service policy do not restore the bundle interface. Service is restored only after configuring the shaper burst size to a value greater than the bundle interface MTU.

Load Balancing

Load balancing function is a forwarding mechanism to distribute traffic over multiple links based on Layer 3 routing information in the router. Per-destination load balancing isonly supported on the router, where the router is allowed to distribute packets over one of the links in the bundle. When the per-destination load

balancing is enabled, all packets for a certain source-destination pair goes through the same link, though there are multiple links available. In other words, per-destination load balancing can ensure that packets for a certain source-destination pair could arrive in order.

Layer 3 Load Balancing on Link Bundles

Layer 3 load balancing for link bundles is done on Ethernet Flow Points (EFPs) and is based on the IPv4 source and destination addresses in the packet. When Layer 3 service-specific load balancing is configured, all egress bundles are load balanced based on the IPv4 source and destination addresses. When packets do not have IPv4 addresses, default load-balancing (based on the MAC SA/DA fields in the packet header) is used.

Configure QoS on Link Bundles

QoS is configured on link bundles in the same way that it is configured on individual interfaces.

Guidelines

- When a QoS policy is applied on a bundle in the egress direction, it's also applied at each member interface.
- When a QoS policy is applied on a bundle (ingress direction), it's replicated at each NPU core.
- If a QoS policy is not applied to a bundle interface, both the ingress and egress traffic use the default queue of the per link member port.
- The shape rate that is specified in the bundle policy-map is not an aggregate for all bundle members. The shape rate applied to the bundle depends on the load balancing of the links. For example, if a policy map with a shape rate of 10 Mbps is applied to a bundle with two member links, and if the traffic is always load-balanced to the same member link, then an overall rate of 10 Mbps applies to the bundle. However, if the traffic is load-balanced evenly between the two links, the overall shape rate for the bundle becomes 20 Mbps.
- If a member is deleted from a bundle, the total bundle statistics changes because the statistics that belongs to the detached link is lost.
- The QoS policy that is applied on bundle is inherited to all its member links and the reference bandwidth that is used to calculate shaper/bandwidth is applied as per the physical member interface bandwidth, and not the bundle as a whole.

Configuration Example

You have to accomplish the following to complete the QoS configuration on link bundles:

Note The policy works only if it is applied on the ingress direction. The egress is supported on COS, DEI and MPLS exp marking. So the below policy may not work when it is applied on egress.

1. Creating a class-map
2. Creating a policy-map and specifying the respective class-map

3. Specifying the action type for the traffic

   Refer Attach a Traffic Policy to an Interface, on page 16 for details on step 1, 2 and 3.

4. Creating a link bundle
5. Applying traffic policy to the link bundle

```
/* Configure Ether-Bundle and apply traffic policy */
Router(config)# interface Bundle-Ether 12000
Router(config-if)# mtu 9100
Router(config-if)# service-policy input ingress
Router(config-if)# service-policy output egress
Router(config-if)# ipv4 address 100.12.0.0 255.255.255.254
Router(config-if)# bundle maximum-active links 64
Router(config-if)# commit
```

Running Configuration

This example shows how a traffic policy is applied on an Ethernet link bundle. The policy is applied to all interfaces that are members of the Ethernet link bundle.

```
/* Policy-map */

policy-map ingress
 class inet4-classifier-af1
  set qos-group 1
 !
 class inet4-classifier-af2
  set qos-group 2
 !
 class inet4-classifier-af3
  set qos-group 3
 !
 class inet4-classifier-af4
  set qos-group 4
 !
 class inet4-classifier-be1
  set qos-group 5
 !
 class inet4-classifier-nc1
  set qos-group 6
 !
 class class-default
 !
 end-policy-map
!

/* Ether Bundle */
interface Bundle-Ether12000
 mtu 9100
 service-policy input ingress
 service-policy output egress
 ipv4 address 100.12.0.0 255.255.255.254
 load-interval 30
 flow ipv4 monitor FMM-V4 sampler SM ingress
 flow ipv6 monitor FMM-V6 sampler SM ingress
 flow mpls monitor FMM-MPLS sampler SM ingress
 ipv4 access-group IPV4ACL_101 ingress
 ipv6 access-group IPV6ACL_101 ingress
```

```
!
```

Verification

- Verify that the bundle status is UP.

```
router# show bundle bundle-ether 1200
Wed Dec 16 19:55:49.974 PST

Bundle-Ether12000
  Status:                                  Up
  Local links <active/standby/configured>:   35 / 0 / 35
  Local bandwidth <effective/available>:     3500000000 (3500000000) kbps
  MAC address (source):                    ea3b.745f.c4b0 (Chassis pool)
  Inter-chassis link:                      No
  Minimum active links / bandwidth:        1 / 1 kbps
  Maximum active links:                    64
  Wait while timer:                        2000 ms
  Load balancing:                          Default
  LACP:                                    Operational
    Flap suppression timer:                Off
    Cisco extensions:                      Disabled
    Non-revertive:                         Disabled
  mLACP:                                   Not configured
  IPv4 BFD:                                Not configured

  Port                  Device           State        Port ID         B/W, kbps
  --------------------  ---------------  -----------  --------------  ----------
  Hu0/4/0/0             Local            Active       0x8000, 0x0009   100000000
      Link is Active
  Hu0/4/0/1             Local            Active       0x8000, 0x000a   100000000
      Link is Active
- - -
- - -
  Hu0/4/0/35            Local            Active       0x8000, 0x002b   100000000
      Link is Active
```

- Verify the bundle statistics:

```
router# show policy-map interface bundle-ether 12000

Bundle-Ether12000 input: ingress

Class inet4-classifier-af1
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :          4647401962/21236124455654       26403040
    Transmitted         :          4647401962/21236124455654       26403040
    Total Dropped       :                   0/0                    0
Class inet4-classifier-af2
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :          4502980177/20576584333939       25571493
    Transmitted         :          4502980177/20576584333939       25571493
    Total Dropped       :                   0/0                    0
Class inet4-classifier-af3
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :          4647404125/21236213667880       26389086
    Transmitted         :          4647404125/21236213667880       26389086
    Total Dropped       :                   0/0                    0
Class inet4-classifier-af4
  Classification statistics          (packets/bytes)     (rate - kbps)
```

```
    Matched             :          9291188840/42456120548683       52771168
    Transmitted         :          9291188840/42456120548683       52771168
    Total Dropped       :                   0/0                    0
Class inet4-classifier-be1
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :          4647413429/21235847852686       26393414
    Transmitted         :          4647413429/21235847852686       26393414
    Total Dropped       :                   0/0                    0
Class inet4-classifier-nc1
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :          9294887621/42473100149807       52778258
    Transmitted         :          9294887621/42473100149807       52778258
    Total Dropped       :                   0/0                    0

Class class-default
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :                 0/0                      0
    Transmitted         :                 0/0                      0
    Total Dropped       :                 0/0                      0

Bundle-Ether12000 output: egress

Class c1
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :         16665494532/75878118942463       8760591
    Transmitted         :         16655834643/75834136022017       8760591
    Total Dropped       :             9659889/43982920446          0
  Queueing statistics
    Queue ID                             : None (Bundle)
    Taildropped(packets/bytes)           : 9659889/43982920446
Class c2
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :         16665421959/75877849543188       8718687
    Transmitted         :         16665421959/75877849543188       8718687
    Total Dropped       :                   0/0                    0
  Queueing statistics
    Queue ID                             : None (Bundle)
    Taildropped(packets/bytes)           : 0/0
Class c3
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :         16665247833/75877509455458       8703470
    Transmitted         :         16665187414/75877234624197       8703470
    Total Dropped       :               60419/274831261            0
  Queueing statistics
    Queue ID                             : None (Bundle)
    Taildropped(packets/bytes)           : 60419/274831261
Class c4
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :         33330896131/151755393012945      17470745
    Transmitted         :         33330745421/151754709368565      17470745
    Total Dropped       :              150710/683644380            0
  Queueing statistics
    Queue ID                             : None (Bundle)
    Taildropped(packets/bytes)           : 150710/683644380
Class c5
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :         16878910340/76849791869834       8833394
    Transmitted         :         16878849464/76849514633309       8833394
    Total Dropped       :               60876/277236525            0
  Queueing statistics
    Queue ID                             : None (Bundle)
    Taildropped(packets/bytes)           : 60876/277236525
Class c6
  Classification statistics          (packets/bytes)     (rate - kbps)
```

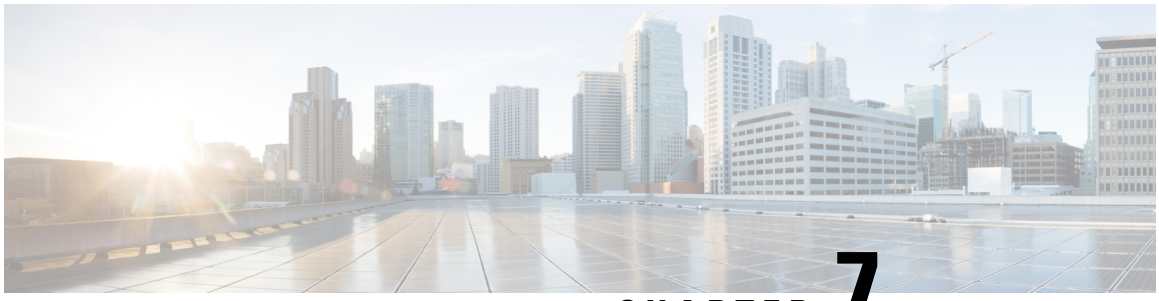
```
    Matched             :         33330898844/151756094112925      17456785
    Transmitted         :         33330752668/151755427708382      17456785
    Total Dropped       :              146176/666404543            0
  Queueing statistics
    Queue ID                             : None (Bundle)
    Taildropped(packets/bytes)           : 146176/666404543
Class c7
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :              244106/79922040             74
    Transmitted         :              244106/79922040             74
    Total Dropped       :                   0/0                    0
  Queueing statistics
    Queue ID                             : None (Bundle)
    Taildropped(packets/bytes)           : 0/0
Class class-default
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :        267075066180/1215993441123215     139917482
    Transmitted         :        267075066180/1215993441123215     139917482
    Total Dropped       :                   0/0                    0
  Queueing statistics
    Queue ID                             : None (Bundle)
    Taildropped(packets/bytes)           : 0/0
```

Related Topics

- QoS on Link Bundles, on page 81

Associated Commands

- bundle maximu-active links
- interface Bundle-Ether



CHAPTER 7

Configuring Hierarchical Modular QoS

Hierarchical QoS (H-QoS) is a QoS model that enables you to specify QoS behavior at multiple levels of hierarchy. This chapter provides information about this feature and the different steps involved in configuring it.

Table 12: Feature History for Hierarchical Modular QoS

Release	Modification
Release 6.3.1	Initial Release

This chapter covers the following topics:

- Overview of Hierarchical Modular QoS, on page 87
- Restrictions, Limitations, and Requirements for Configuring H-QoS, on page 87
- Configuring Hierarchical Queuing, on page 89

Overview of Hierarchical Modular QoS

Hierarchical QoS (H-QoS) allows you to specify QoS behavior at multiple policy levels, which provides a high degree of granularity in traffic management.

H-QoS is applied on the router interface using nested traffic policies. The first level of traffic policy, the parent traffic policy, is used for controlling the traffic at the main interface or sub-interface level. The second level of traffic policy, the child traffic policy, is used for more control over a specific traffic stream or class. The child traffic policy, is a previously defined traffic policy, that is referenced within the parent traffic policy using the **service-policy** command.

Two-level H-QoS is supported on both ingress and egress directions on all line cards and on physical or bundle main interfaces and sub-interfaces.

Restrictions, Limitations, and Requirements for Configuring H-QoS

There are some key restrictions, limitations, and requirements for configuring H-QoS on your router to ensure optimal performance and compliance.

Parent Traffic Policy Restrictions

- The parent traffic policy only supports the traffic class of type class-default.
- The parent traffic policy only supports the class-action shape, and no other queuing action can be configured in it.
- For congestion avoidance and management, the traffic shaper in the parent traffic policy calculates the queue limit and drop priority

Child Policy Requirements

- While configuring on the router, it is mandatory that the priority class must have a traffic shaper in the child traffic policy.
- The sum of the bandwidth of the child policies must be less than the parent policy's traffic shaper.

H-QoS Profile Limitations

- H-QoS profile and ingress peering profile don't work simultaneously. Hence, features requiring a peering profile also do not work with the H-QoS profile enabled.
- The PBTS feature does not work when the H-QoS profile is enabled. This is due to TCAM limitations.
- A maximum of 896 bundle subinterfaces are supported in the system, even if there are no QoS policies applied. This is due to an internal LAG_ID resource consumption in H-QoS profile mode for bundle subinterfaces with or without QoS policies being applied.
- A maximum of 4 priority levels are supported in H-QoS profile mode, unlike the default mode where 7-priority levels are supported. The restriction also applies to physical and bundle main interface policies where 7-level priorities were previously used in non-H-QoS profile mode.

Bandwidth Configuration Limitations

- Bandwidth and Bandwidth remaining configurations are not supported simultaneously within the same policy-map. If a class has bandwidth (CIR), other classes must also have only bandwidth configuration. If a class-map has bandwidth remaining percent/ratio (EIR), other classes should also have only the bandwidth remaining configuration. Shaping is applied on any class.
- In H-QoS mode, if multiple queues are configured with BRR and there is high congestion in the LP (low-priority) queues, one BRR queue receives more credits than the others. The deviation is proportional to the congestion in the LP queues.
- The granularity of bandwidth or the bandwidth remaining ratio (BRR) is 1:64, compared to 1:4096 in non-H-QoS mode. Consequently, there may be differences in bandwidth performance accuracy depending on the values used.

Priority Classes Requirements

Priority classes must be rate-limited using a shaping configuration. The effective shaper value is used as the priority bandwidth reservation. The total priority bandwidth reservations across all subinterfaces and main interfaces must not exceed the network interface (NIF) port speed to prevent oversubscription of priority traffic. Non-priority class rates and parent shaping can be oversubscribed.

Multicast Traffic Restriction

Filtering egress IPv4 and IPv6 multicast traffic is not supported when H-QoS is configured on the router.

Configuring Hierarchical Queuing

Before you configure H-QoS, you must enable the H-QoS profile on the router. After enabling H-QoS profile, reload the router, as shown in the following configuration.

```
admin
hw-module location all reload
Router# configure
Router(config)# hw-module profile qos hqos-enable
Router(config)# commit
Router# admin
sysadmin-vm:0_RP0# hw-module location all reload
```

The steps that are involved in configuring hierarchical queuing are as follows:

1. Configure a class-map.
2. Configure a child traffic policy using the class-map that was configured in the previous step.
3. Configure a parent traffic policy and add the child traffic policy in it.

The parent traffic policy is the H-QoS traffic policy and it can be applied on physical or bundle main interfaces and sub-interfaces.

Configuration Example

Configuration of a class-map is as follows:

```
Router# configure
Router(config)# class-map match-any tc2
Router(config-cmap)# match traffic-class 1
Router(config-cmap)# end-class-map
Router(config)# commit
```

Configuration of a child traffic policy is as follows:

```
Router# configure
Router(config)# policy-map child
Router(config-pmap)# class tc2
Router(config-pmap-c)# shape average percent 20
Router(config-pmap-c)# exit
Router(config-pmap)# class class-default
Router(config-pmap-c)# shape average percent 1
Router(config-pmap)# end-policy-map
Router(config)# commit
```

Configuration of a parent traffic policy is as follows:

```
Router# configure
Router(config)# policy-map parent
Router(config-pmap)# class class-default
Router(config-pmap-c)# service-policy child
Router(config-pmap-c)# shape average percent 50
```

```
Router(config-pmap)# end-policy-map
Router(config)# commit
```

Running Configuration

```
/* Configuration of a Class-map */
class-map match-any tc2
 match traffic-class 1
 end-class-map
!
/* Configuration of a Child Traffic Policy */
policy-map child
 class tc2
  shape average percent 20
 !
 class class-default
  shape average percent 1
 !
 end-policy-map
!
/* Configuration of a Parent Traffic Policy */
policy-map parent
 class class-default
  service-policy child
  shape average percent 50
 !
 end-policy-map
!
```

Applying the Parent Traffic Policy on a Main Interface

```
Router# configure
Router(config)# Interface TenGigE 0/0/0/10
Router(config-int)# service-policy output parent
Router(config-int)# commit
```

Applying the Parent Traffic Policy on a Sub-interface

```
Router# configure
Router(config)# Interface TenGigE 0/0/0/10.1
Router(config-int)# service-policy output parent
Router(config-int)# commit
```

Verification

Verify if the H-QoS traffic policy is applied correctly on the interface using the commands **show qos interface** *interface-name* **output**. In the following example, the **Level1 Class** gives information about the class-map that is associated with the parent traffic policy and the **Level2 Class** gives information about the class-maps that are associated with the child traffic policy.

```
RP/0/RP0/CPU0:ios#show qos interface ten0/0/0/10 output

NOTE:- Configured values are displayed within parentheses
Interface TenGigE0/0/0/10 ifh 0x1e0  -- output policy
NPU Id:                        0
Total number of classes:       3
Interface Bandwidth:           10000000 kbps
VOQ Base:                      1136
Accounting Type:               Layer1 (Include Layer 1 encapsulation and above)
```

```
------------------------------------------------------------------------------
Level1 Class                                =   class-default
Queue Max. BW.                              =   no max (50 %)
Queue Min. BW.                              =   0 kbps (default)
Inverse Weight / Weight                     =   0 / (BWR not configured)
   Level2 Class                                =   tc2
   Egressq Queue ID                            =   1138 (LP queue)
   Queue Max. BW.                              =   1020015 kbps (20 %)
   Queue Min. BW.                              =   0 kbps (default)
   Inverse Weight / Weight                     =   1 / (BWR not configured)
   Guaranteed service rate                     =   1000000 kbps
   TailDrop Threshold                          =   1253376 bytes / 10 ms (default)
   WRED not configured for this class
   Level2 Class                                =   class-default
   Egressq Queue ID                            =   1136 (Default LP queue)
   Queue Max. BW.                              =   50625 kbps (1 %)
   Queue Min. BW.                              =   0 kbps (default)
   Inverse Weight / Weight                     =   1 / (BWR not configured)
   Guaranteed service rate                     =   50000 kbps
   TailDrop Threshold                          =   62720 bytes / 10 ms (default)
   WRED not configured for this class
```

The statistics for the packets that have matched the different traffic classes of the parent and child traffic policies can be viewed using the command **show policy-map interface** *interface-name* **output**. Also, this command also shows the number of packets that are transmitted or dropped when the specified action is applied on the packets that have matched the respective traffic class.

```
Router# show policy-map interface ten0/0/0/10 output

TenGigE0/0/0/10 output: parent
Class class-default
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :          2313578823/296138089344          8494665
    Transmitted         :           232805738/29799134464           854465
    Total Dropped       :          2080773085/266338954880          7640200
  Policy child Class tc2
    Classification statistics          (packets/bytes)     (rate - kbps)
      Matched             :          2313578823/296138089344          8494665
      Transmitted         :           232805738/29799134464           854465
      Total Dropped       :          2080773085/266338954880          7640200
    Queueing statistics
      Queue ID                               : 1138

      Taildropped(packets/bytes)             : 2080773085/266338954880
  Policy child Class class-default
    Classification statistics          (packets/bytes)     (rate - kbps)
      Matched             :                    0/0                    0
      Transmitted         :                    0/0                    0
      Total Dropped       :                    0/0                    0
    Queueing statistics
      Queue ID                               : 1136

      Taildropped(packets/bytes)             : 0/0
```

When using hierarchical policers, there is no independent set of hardware counters to store the parent policer statistics. Instead, parent policer statistics are manipulated in the software to be the sum of all child policers under the same policy-map.

This is shown in the following example where two streams of traffic, with CoS value of 1 and 2 are sent at a speed of 3.5 Gbps each.

```
/*Hierarchical Policy Map Configuration*/
==================================================
```

```
Router# show running-config policy-map Hingress
policy-map Hingress
 class class-default
  service-policy ingress
  police rate 5 gbps peak-rate 9 gbps
  !
 !
 end-policy-map
!
 /*Ingress Policy Map Configuration*/
====================================
Router#show running-config policy-map ingress
policy-map ingress
 class cos1
  set traffic-class 1
  police rate 5 gbps
  !
 !
 class cos2
  set traffic-class 2
  police rate 5 gbps
  !
 !
 class class-default
 !
 end-policy-map
!
 /*Policy Map applied at TenGigE0/0/0/6.100 Interface*/
=======================================================
Router#show policy-map interface tenGigE 0/0/0/6.100 input

TenGigE0/0/0/6.100 input: Hingress

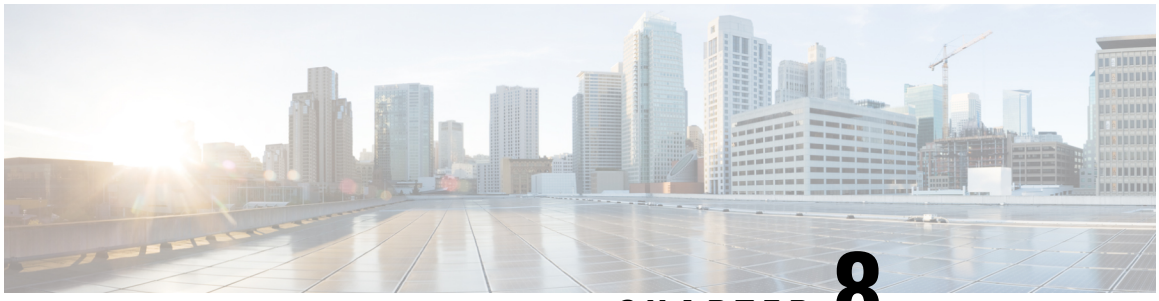
Class class-default
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :           856717937/109659895936         6683676
    Transmitted         :           856717937/109659895936         6683676
    Total Dropped       :                   0/0                    0
  Policing statistics                (packets/bytes)     (rate - kbps)
    Policed(conform)    :           856717937/109659895936         6683674
    Policed(exceed)     :                   0/0                    0
    Policed(violate)    :                   0/0                    0
    Policed and dropped :                   0/0

  Policy ingress Class cos1
    Classification statistics          (packets/bytes)     (rate - kbps)
      Matched             :           437826303/56041766784          3341838
      Transmitted         :           437826303/56041766784          3341838
      Total Dropped       :                   0/0                    0
    Policing statistics                (packets/bytes)     (rate - kbps)
      Policed(conform)    :           437826303/56041766784          3341838
      Policed(exceed)     :                   0/0                    0
      Policed(violate)    :                   0/0                    0
      Policed and dropped :                   0/0
      Policed and dropped(parent policer)  : 0/0

  Policy ingress Class cos2
    Classification statistics          (packets/bytes)     (rate - kbps)
      Matched             :           418891634/53618129152          3341838
      Transmitted         :           418891634/53618129152          3341838
      Total Dropped       :                   0/0                    0
    Policing statistics                (packets/bytes)     (rate - kbps)
      Policed(conform)    :           418891634/53618129152          3341838
      Policed(exceed)     :                   0/0                    0
```

```
      Policed(violate)    :                  0/0                    0
      Policed and dropped :                  0/0
      Policed and dropped(parent policer)  : 0/0

  Policy ingress Class class-default
    Classification statistics          (packets/bytes)     (rate - kbps)
      Matched             :                  0/0                    0
      Transmitted         :                  0/0                    0
      Total Dropped       :                  0/0                    0
Policy Bag Stats time: 0
Policy Bag Stats time: 0
```



CHAPTER 8

QoS for Bridge-Group Virtual Interfaces

Integrated Routing and Bridging (IRB) provides the ability to route between a bridge group and a routed domain with the help of Bridge-Group Virtual Interface (BVI).

The BVI is a virtual interface within the router that acts like a normal routed interface that does not support bridging, but represents the comparable bridge group to routed interfaces within the router. The interface number of the BVI is the number of the bridge group that the virtual interface represents. The number is the link between the BVI and the bridge group.

For more information on IRB/ BVI, please refer the *Interface and Hardware Component Configuration Guide for Cisco NCS 540 Series Routers*

- Information on Qos on BVI, on page 95
- Restrictions on BVI , on page 95
- Classification and Marking , on page 96
- Configuring QoS on BVI , on page 97
- Verifying QoS on BVI, on page 98

Information on Qos on BVI

A BVI integrates Layer2 domain with Layer3 domain by creating a virtual interface in between them. The traffic flow supported for QoS is from the bridged to routed interface.

- A BVI can have bridge domain members from different linecards or NPU or Core.
- A BVI service-policy is applied on all linecards and within each linecard it is replicated per NPU. If any policers are configured in the qos policy map is applied to the BVI interface, then the policer is configured on each core of the NPU, and is shared among all the interfaces on that NPU core.

Restrictions on BVI

- Egress QoS on BVI is not supported for Layer2 flows for routed to bridge domain traffic.
- QoS on BVI is not supported for:
  - Queuing (priority, bandwidth, bandwidth remaining, sShaping, queue-limit, random-detect)
  - Shared policy instance

- Overhead accounting
- Percentage policer at the parent level

- Conditional EXP marking is only supported for Layer3 VPN when BVI is the AC interface or the packets have destination MAC address as BVI interface.

  Conditional EXP marking is not supported for EVPN (Layer2 traffic – non-BVI destination MAC traffic).

- Inheriting a policy map must be from either from a BVI interface or main interface.

  The policy map on a bridge domain sub-interface is inherited by the main interface and BVI interface. If the policy is applied on both the interfaces, then the policy is inherited on the last applied interface. This configuration is not supported and should be avoided to undesired or unknown behavior.

- A policy on a BVI interface should be applied after the BVI is added to the bridge domain as a routed interface. If the policy is applied before, then the policy has no effect.
- The show QoS interface command is not supported for the BVI interface.

Classification and Marking

The following features are supported:

- Classification
- Policing (level 1 and level 2)
- Ingress marking

Table 13: Classification and Marking

	Classification	Marking	Direction	Comment
Qos-group	No	Yes	Ingress	Used for Egress marking policy
Discard-class	No	Yes	Ingress	Used for VOQ selection and Egress queuing policy
Precedence	Yes	Yes	Ingress	
DSCP	Yes	Yes	Ingress	
Vlan	No	No		
CoS	Yes	No	Ingress	Only for Layer2 flows (bridge member).
Dei	Yes	No	Ingress	Only for Layer2 flows (bridge member).

	Classification	Marking	Direction	Comment
EXP	No	Yes	Ingress	

Configuring QoS on BVI

```
interface TenGigE0/0/0/0
 l2transport
 !
!

interface TenGigE0/0/0/1
 l2transport
 !
!

interface BVI1
 service-policy input bvi-ingress
 ipv4 address 1.1.1.1 255.255.255.0
!

l2vpn
 bridge group bg1
  bridge-domain bd1
   interface TenGigE0/0/0/0
   !
   interface TenGigE0/0/0/1
   !
   routed interface BVI1
   !
  !
 !
!

class-map match-any prec1
 match precedence 1
 end-class-map
!



class-map match-any dscp_af22
 match dscp af22
 end-class-map
!

policy-map bvi-ingress
 class prec1
  set traffic-class 1
  set mpls experimental imposition 3
  police rate 100 mbps
  !
 !
 class dscp_af22
  set traffic-class 2
  set mpls experimental imposition 4
  police rate 200 mbps
  !
 !
 class class-default
```

```
 !
 end-policy-map
!
```

Verifying QoS on BVI

Use the **show policy-map interface input** command to collect statistics from all linecards.

```
Router# show policy-map interface bvI 1 input

BVI1 input: bvi-ingress

Class prec1
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :                  0/0                   0
    Transmitted         :                  0/0                   0
    Total Dropped       :                  0/0                   0
Class dscp_af22
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :                  0/0                   0
    Transmitted         :                  0/0                   0
    Total Dropped       :                  0/0                   0
Class class-default
  Classification statistics          (packets/bytes)     (rate - kbps)
    Matched             :                  0/0                   0
    Transmitted         :                  0/0                   0
    Total Dropped       :                  0/0                   0
```