Congestion Avoidance Commands

Note All commands applicable for the Cisco NCS 5500 Series Router are also supported on the Cisco NCS 540 Series Router that is introduced from Cisco IOS XR Release 6.3.2. References to earlier releases in Command History tables apply to only the Cisco NCS 5500 Series Router.

Note

- Starting with Cisco IOS XR Release 6.6.25, all commands applicable for the Cisco NCS 5500 Series Router are also supported on the Cisco NCS 560 Series Routers.
- Starting with Cisco IOS XR Release 6.3.2, all commands applicable for the Cisco NCS 5500 Series Router are also supported on the Cisco NCS 540 Series Router.
- References to releases before Cisco IOS XR Release 6.3.2 apply to only the Cisco NCS 5500 Series Router.
- Cisco IOS XR Software Release 7.0.1 specific updates are not applicable for the following variants of Cisco NCS 540 Series Routers:
  - N540-28Z4C-SYS-A
  - N540-28Z4C-SYS-D
  - N540X-16Z4G8Q2C-A
  - N540X-16Z4G8Q2C-D
  - N540X-16Z8Q2C-D
  - N540-12Z20G-SYS-A
  - N540-12Z20G-SYS-D
  - N540X-12Z16G-SYS-A
  - N540X-12Z16G-SYS-D

This chapter describes commands used to avoid congestion.

Congestion avoidance is achieved through packet dropping.

bandwidth (QoS)

To specify the minimum bandwidth allocated to a class belonging to a policy map, use the **bandwidth** command in policy map class configuration mode. To remove the bandwidth specified for a class, use the **no** form of this command.

bandwidth {*rate* [*units*] | **percent** *percentage-value*}
no bandwidth {*rate* [*units*] | **percent** *percentage-value*}

Syntax Description

rate	Minimum bandwidth, in the units specified, to be assigned to the class. Range is from 1 to 4294967295.
units	Specifies the units for the bandwidth. Values can be: • **bps**—bits per second • **gbps**—gigabits per second • **kbps**—kilobits per second (default) • **mbps**—megabits per second
percent *percentage-value*	Specifies the amount of minimum guaranteed bandwidth, based on an absolute percentage of available bandwidth. Range is from 1 to 100.

Command Default

The default units is kbps.

Command Modes

Policy map class configuration

Command History

Release	Modification
Release 6.0	This command was introduced.

Usage Guidelines

The **bandwidth** command is used to specify the minimum guaranteed bandwidth allocated for traffic matching a particular class. Bandwidth may be defined as a specific value or may be set as a percentage of the interface bandwidth.

If a percentage value is set, the accuracy that can be expected is 1 percent.

The **bandwidth** command is supported only in the egress direction.

A policy map can have a single bandwidth statement per class. Both percentage and actual value bandwidth configurations can be used within a policy map.

The **bandwidth** command does not specify how the bandwidth is to be shared. Instead it specifies how much bandwidth is guaranteed per class, by setting the number of tokens that are assigned to the token bucket of a particular class. For configured behavior to work correctly, you must ensure that the sum of the bandwidths plus any priority traffic is not greater than the bandwidth of the interface itself. If the interface is oversubscribed, unpredictable behavior results.

Task ID

Task ID	Operations
qos	read, write

Examples

This example shows how to guarantee 50 percent of the interface bandwidth to a class called class1 and 10 percent of the interface bandwidth to a class called class2:

```
RP/0/RP0/CPU0:router(config)# policy-map policy1
RP/0/RP0/CPU0:router(config-pmap)# class class1
RP/0/RP0/CPU0:router(config-pmap-c)# bandwidth percent 50
RP/0/RP0/CPU0:router(config-pmap-c)# exit
RP/0/RP0/CPU0:router(config-pmap)# class class2
RP/0/RP0/CPU0:router(config-pmap-c)# bandwidth percent 10
```

bandwidth remaining

To specify how to allocate leftover bandwidth to various classes, use the **bandwidth remaining** command in policy map class configuration mode. To return to the system defaults, use the **no** form of this command.

bandwidth remaining [**percent** *percentage-value* | **ratio** *ratio-value*]
no bandwidth remaining [**percent** *percentage-value* | **ratio** *ratio-value*]

Syntax Description

percent *percentage-value*	Specifies the amount of guaranteed bandwidth, based on an absolute percentage of the available bandwidth. Range is from 1 to 100.
ratio *ratio-value*	Specifies the amount of guaranteed bandwidth, based on a bandwidth ratio value. Range is 1 to 2000.

Command Default

No bandwidth is specified.

Command Modes

Policy map class configuration

Command History

Release	Modification
Release 6.0	This command was introduced.

Usage Guidelines

Bandwidth, bandwidth remaining, shaping, queue-limit and WRED commands can be configured together in the same class.

Note The **bandwidth remaining** command is supported only in the egress direction.

The available bandwidth is equally distributed among those queueing classes that do not have the remaining bandwidth explicitly configured.

The **bandwidth remaining** command is used to proportionally allocate bandwidth to the particular classes, but there is no reserved bandwidth capacity.

On egress, if the **bandwidth remaining** command is not present, then the bandwidth is shared equally among the configured queueing classes present in the policy-map.

Task ID

Task ID	Operations
qos	read, write

Examples

This example shows how the remaining bandwidth is shared by classes class1 and class2 in a 20:80 ratio.

```
RP/0/RP0/CPU0:router(config)# policy-map policy1
```

```
RP/0/RP0/CPU0:router(config-pmap)# class class1
RP/0/RP0/CPU0:router(config-pmap-c)# bandwidth remaining percent 20
RP/0/RP0/CPU0:router(config-pmap-c)# exit
RP/0/RP0/CPU0:router(config-pmap)# class class2
RP/0/RP0/CPU0:router(config-pmap-c)# bandwidth remaining percent 80
```

hw-module profile priority-flow-control

To configure PFC threshold values per line card, use the **hw-module profile priority-flow-control** command in XR Config mode. To return to the system defaults, use the **no** form of this command.

Note For details on the show controller, set controller, clear controller and priority-flow-control watchdog commands, see the *Interface and Hardware Component Command Reference for Cisco NCS 5500 and NCS 540 and NCS 560 Series Routers*.

To disable this feature, use the no form of this command.

hw-module profile priority-flow-control { { | **location** <*lc*> } | { **traffic-class** *number* | | | **pause-threshold** *x-off size units* | | **resume-threshold** *x-on size units* | **headroom** *value units* | | } }

Syntax Description

location	<*lc*> —Line card location
traffic-class	Traffic class to configure PFC threshold values.
traffic class number	0—7 **Note** Reload the line card for all traffic-class addition and deletions. Parameters within an already configured traffic-class can be edited 'in place' without requiring a line card reload.
pause-threshold	Specifies the buffer limit at which the flow is paused.
x-off size	Buffer limit for pausing flow.
units	Unit of measurement for the x-off size in bytes.
resume-threshold	Specifies the buffer limit at which the flow is resumed.
x-on size	Buffer limit for resuming flow.
units	Unit of measurement for the x-on size in bytes.
headroom	Specifies the headroom.
value	Headroom value
units	Unit of measurement for the headroom value in bytes.

Command Default

No default behavior or values.

Command Modes

XR Config

Command History

Release	Modification
Release 6.6.4	This command was introduced and replaced the pause, on page 11 command.

Usage Guidelines

If you add a new traffic class and configure PFC threshold values for the first time on that traffic class, you must reload the line card.

Task ID

Task ID	Operations
qos	read, write

Examples

This example shows how to configure the PFC threshold values for traffic-class 3 and traffic-class 4.

```
RP/0/RP0/CPU0:router(config)#hw-module profile priority-flow-control location 0/0/CPU0
RP/0/RP0/CPU0:router(config-pfc-loc)#traffic-class 3 pause-threshold 403200 bytes
resume-threshold 40320 bytes headroom 441600 bytes
RP/0/RP0/CPU0:router(config-pfc-loc)#traffic-class 4 pause-threshold 403200 bytes
resume-threshold 40320 bytes headroom 441600 bytes
RP/0/RP0/CPU0:router(config-pfc-loc)#exit
RP/0/RP0/CPU0:router(config-pfc-loc)#commit
```

hw-module profile qos ecn-marking-stats

To enable the display of counters for ECN-marked and transmitted packets, configure the

hw-module profile qos ecn-marking-stats command in the XR Config mode. To disable this feature, use the no form of this command.

hw-module profile qos ecn-marking-stats

Syntax Description

This command has no keywords or arguments.

Command Default

The display of counters for ECN-marked and transmitted packets is disabled by default.

Command Modes

XR Config mode

Command History

Release	Modification
Release 7.5.4	This command was introduced.

Usage Guidelines

You must reload the affected line card to enable the display of counters for ECN-marked and transmitted packets.

Task ID

Task ID	Operation
qos	read, write

The following example shows how to enable the display of counters for ECN-marked and transmitted packets.

```
Router#config
Router(config)#hw-module profile qos ecn-marking-stats
In order to activate this profile, you must manually reload the chassis/all line cards
Router(config)#commit
```

hw-module profile qos wred-stats-enable

To enable the display of Weighted Random Early Detection (WRED) statistics per class, use the

hw-module profile qos wred-stats command in the XR Config mode. To disable this feature, use the no form of this command.

This command has no keywords or arguments.

Command Default

The WRED statistics mode is disabled by default.

Command Modes

XR Config mode

Command History

Release	Modification
Release 7.4.1	This command was introduced.

Usage Guidelines

You must reload the affected line card to enable the WRED statistics mode.

Task ID

Task ID	Operation
qos	read, write

The following example shows how to enable the wred-stats mode.

```
RP/0/RP0/CPU0:router#config
RP/0/RP0/CPU0:router(config)#hw-module profile qos wred-stats-enable
RP/0/RP0/CPU0:router(config)#commit
RP/0/RP0/CPU0:router(config)#exit
RP/0/RP0/CPU0:router#reload
```

pause

This configuration allows you to configure nondefault Priority Flow Control (PFC) buffer thresholds per priority queue through an 8-priority egress queuing policy applied on a PFC-enabled interface.

Note From Release 6.6.4 onwards, the **pause** command is deprecated. To configure PFC thresholds, see hw-module profile priority-flow-control, on page 7.

The router ignores the **buffer-size parameter** , although the configuration is not optional. Hence, it is recommended that you keep the buffer-size on par with the queue-limit.

pause {**buffer-size***units* | | | **pause-threshold***x-off size units* | | **resume-threshold***x-on size units*}

Syntax Description

buffer-size	Buffer size for ingress traffic.
units	(Optional) Unit of measurement for the buffer size. Values can be: **bytes—bytes** (default) **gbytes—gigabytes** **kbytes—kilobytes** **mbytes—megabytes**
pause-threshold	Specifies the buffer limit at which the flow is paused.
x-off size	Buffer limit for pausing flow.
units	(Optional) Unit of measurement for the x-off size. Values can be: **bytes—bytes** (default) **gbytes—gigabytes** **kbytes—kilobytes** **mbytes—megabytes**
resume-threshold	Specifies the buffer limit at which the flow is resumed.
x-on size	Buffer limit for resuming flow.
units	(Optional) Unit of measurement for the x-off size. Values can be: **bytes—bytes** (default) **gbytes—gigabytes** **kbytes—kilobytes** **mbytes—megabytes**

Command Default

When you do not configure pause and resume thresholds for a priority queue, the pause threshold (*x-off size*) is set to 128 mbytes and the resume threshold (*x-on size*) is set to 8 kbytes.

Command Modes

Policy map class configuration.

Command History

Release	Modification
Release 6.6.3	This command was introduced.
Release 6.6.4	This command was deprecated and replaced by the hw-module profile priority-flow-control, on page 7 command.

Usage Guidelines

To configure PFC thresholds, see hw-module profile priority-flow-control, on page 7.

Task ID

Task ID	Operations
Qos	read, write

Examples

In this example, the policy map is configured with pause parameter.

```
RP/0/RP0/CPU0:router(config)# #policy-map pmap_out
RP/0/RP0/CPU0:router(config-pmap)# #class tc1
RP/0/RP0/CPU0:router(config-pmap-c)# #pause ?

 buffer-size  Configure buffer size

RP/0/RP0/CPU0:router(config-pmap-c)# pause buffer-size 100 kbytes ?

pause-threshold  Configure pause threshold

RP/0/RP0/CPU0:router(config-pmap-c)# )#$ pause-threshold 100?

  bytes             Bytes
  gbytes            Gigabytes
  kbytes            Kilobytes
  mbytes            Megabytes

RP/0/RP0/CPU0:router(config-pmap-c)# )#$ resume-threshold 90

  bytes             Bytes
  gbytes            Gigabytes
  kbytes            Kilobytes
  mbytes            Megabytes
```

queue-limit

To specify or modify the maximum number of packets the queue can hold for a class policy configured in a policy map for each port, use the **queue-limit** command in policy map class configuration mode. To remove the queue packet limit from a class, use the **no** form of this command.

queue-limit *value* [*unit*] | *percent <1-100>*
no queue-limit

Syntax Description

value	Maximum threshold for tail drop in bytes. Range is from 1 to 4294967295.
unit	(Optional) Units for the queue limit value. Values can be: • **bytes** —bytes • **kbytes** —kilobytes • **mbytes** —megabytes • **ms** —milliseconds • **packets** —packets (default) • **us** —microseconds **Note** When the specified *units* is packets, packets are assumed to be 256 bytes in size.
percent	(Optional) Allows you to specify queue limit thresholds as a percentage of the total buffer limit for each port. This makes your provisioning model simpler and makes it easier for you to adjust the queue burst limit, irrespective of the queue's service rate. The calculation is based on the assumption that a port takes 40 milli-seconds of buffering at port-rate. This option was introduced in Release 6.1.2.

Command Default

The default value is 10 milliseconds for all queues including the high-priority queues.

Command Modes

Policy map class configuration

Command History

Release	Modification
Release 6.0	This command was introduced.

Usage Guidelines

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. The default value is 10 milliseconds for all queues including the high-priority queues.

The **queue-limit** command is supported only in the egress direction.

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

Tail drop is a congestion avoidance technique that drops packets when an output queue is full, until congestion is eliminated.

Use the **show qos interface** command to display the queue limit and other QoS values.

Queue Limit Default Values

These default values are used when **queue-limit** is not configured in the class:

- If QoS is configured and Weighted Random Early Detection (WRED) is not configured:
  - Queue limit is 10 ms at the guaranteed service rate of the queue for non-priority queues.
  - Queue limit is 10 ms at the interface rate for priority classes
- If QoS is configured and WRED is configured:
  - Queue limit is two times the WRED maximum threshold. The maximum threshold can be an explicitly configured value or an implicit 10 ms.
  - If more than one WRED profile is configured in the class, the maximum threshold is the maximum for all profiles.
  - When the **queue-limit** is configured in time units, the guaranteed service rate (for the non-priority class) or the interface rate (for the priority class) is used to compute the queue-limit.

These restrictions apply to queue limits:

- For releases before Release 6.3.2, the queue limit should be at least the maximum MTU size, which is fixed at 9 * 1024 bytes = 9kb. From Release 6.3.2 onwards, the minimum queue limit is the interface MTU (which is dynamically configured).
- Queue limit should be 3 GB, which is the maximum packet buffer size in ingress and egress queuing ASICs.
- Only time-based units are allowed on bundle targets.

Guaranteed Service Rate

The guaranteed service rate is defined as the service rate of the queue when all queues are backlogged and derived as:

minimum_bandwidth + (bandwidth_remaining_percent * unallocated_bandwidth)

This example shows the guaranteed service rate calculation:

```
policy-map sample_policy
  class c1
    bandwidth percent 30
    bandwidth remaining percent 40
  class c2
    bandwidth percent 20
  class class-default
```

guaranteed service rate of c1 = 30 percent LR + (40 percent * 50 percent * LR)

guaranteed service rate of c2 = 20 percent LR + (30 percent * 50 percent * LR)

guaranteed service rate of class-default = 30 percent * 50 percent * LR

- Where LR is line rate of the target on which service policy "sample_policy" is attached.
- 50 percent is unallocated bandwidth.

Task ID

Task ID	Operations
qos	read, write

Examples

This example shows how to set the queue limit for a class to 1000000 packets for policy map policy1:

```
RP/0/RP0/CPU0:router(config)# policy-map policy1
RP/0/RP0/CPU0:router(config-pmap)# class class1
RP/0/RP0/CPU0:router(config-pmap-c)# queue-limit 1000000
```

random-detect

To enable random early detection (RED), use the **random-detect** command in policy map class configuration mode. To remove RED, use the **no** form of this command.

random-detect [**default** | [**discard-class** *value*] | [[**min-threshold** *value*] [**max-threshold** *value*] **probability** *percentage probability value*]]]

no random-detect

Syntax Description

default	Enables RED with default minimum and maximum thresholds.
discard-class *value*	Discard class ID. Prior to IOS XR Release 7.1.1, an integer from 0 through 2, to be marked on the packet. From IOS XR Release 7.1.1, an integer from 0 through 3, to be marked on the packet. (See **Usage Guidelines** below.)
min-threshold	Minimum threshold in number of packets. The value range of this argument is from 0 to 1073741823 in packets.
max-threshold	Maximum threshold in number of packets. The value range of this argument is from the value of the *min-threshold* argument to 1073741823. When the queue length exceeds the maximum threshold, RED drops all packets with the specified discard class value.
units	(Optional) Units for the threshold values. Values can be: • **bytes**—bytes • **gbytes**—gigabytes • **kbytes**—kilobytes • **mbytes** —megabytes • **ms** —milliseconds • **packets** —packets (default) • **us** —microseconds
probability *percentage probability value*	Configure WRED mark probability in percent. The value range of this argument is from 1 through 100. **Note** • The default value is 10 • To see which line cards support this option, see *Configure ECN Maximum Marking Probability* in the *Modular QoS Configuration Guide for Cisco NCS 5500 Series Routers*

Command Default

Default unit for *max-threshold* and *min-threshold* is **packets**.

Command Modes

Policy map class configuration

Command History

Release	Modification
Release 6.0	This command was introduced.
Release 7.3.3	The functionality to configure ECN mark probability in percent was introduced.

Usage Guidelines

The RED congestion avoidance technique takes advantage of the congestion control mechanism of TCP. By randomly dropping packets before 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 the 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.

When time units are used, the guaranteed service rate is used to compute thresholds. The default minimum threshold is 6 ms and the maximum threshold is 10 ms.

When the value of the *units* argument is packets, packets are assumed to be 256 bytes in size.

If you set a discard-class of 3, the packet is dropped at ingress itself.

Weighted Random Early Detection

The following restriction applies to Weighted Random Early Detection (WRED):

- For thresholds in time units, the guaranteed service rate is used to calculate the thresholds in bytes.

For bundles, queue limit and WRED thresholds are supported in time units only.

Task ID

Task ID	Operations
qos	read, write

Examples

This example shows how to enable RED using a minimum threshold value of 1000000 and a maximum threshold value of 2000000:

```
RP/0/RP0/CPU0:router(config)# policy-map policy1
RP/0/RP0/CPU0:router(config-pmap)# class class1
RP/0/RP0/CPU0:router(config-pmap-c)# random-detect 1000000 2000000
```

service-policy (interface)

To attach a policy map to an input interface or output interface to be used as the service policy for that interface, use the **service-policy** command in the appropriate configuration mode. To remove a service policy from an input or output interface, use the **no** form of the command.

service-policy {**input** | **output**} *policy-map*
no service-policy {**input** | **output**} *policy-map*

Syntax Description

input	Attaches the specified policy map to the input interface.
output	Attaches the specified policy map to the output interface.
policy-map	Name of a service policy map (created using the **policy-map** command) to be attached.

Command Default

No service policy is specified.

Command Modes

Interface configuration.

Command History

Release	Modification
Release 6.0	This command was introduced.

Usage Guidelines

You can attach a single policy map to one or more interfaces to specify the service policy for those interfaces. The class policies composing the policy map are then applied to packets that satisfy the class map match criteria for the class. To apply a new policy to an interface, you must remove the previous policy. A new policy cannot replace an existing policy.

Task ID

Task ID	Operations
qos	read, write

Examples

This example shows policy map policy2 applied to HundredGigabitEthernet 0/0/0/1.

```
RP/0/RP0/CPU0:router(config)# class-map class2
RP/0/RP0/CPU0:router(config)# match precedence ipv4 2
RP/0/RP0/CPU0:router(config-cmap)# exit

RP/0/RP0/CPU0:router(config)# policy-map policy2
RP/0/RP0/CPU0:router(config-pmap)# class-map class2
RP/0/RP0/CPU0:router(config-pmap-c)# set precedence 3
RP/0/RP0/CPU0:router(config-pmap)# exit

RP/0/RP0/CPU0:router(config)# HundredGigabitEthernet 0/0/0/1
RP/0/RP0/CPU0:router(config-if)# service-policy input policy2
```

This example shows policy map policy 1 applied to Bundle-Ether interface.

```
RP/0/RP0/CPU0:router(config)# interface Bundle-Ether1
RP/0/RP0/CPU0:router(config-if)# service-policy input policy1
RP/0/RP0/CPU0:router(config-if)# exit
```