

## Quality of Service on the Cisco Catalyst 4500 Supervisor Engine 6-E

This document applies to the Cisco® Catalyst® 4500 Supervisor Engine 6-E and Cisco Catalyst 4900M Switch and is based on Cisco IOS® Software Release 12.2(50)SG.

### Introduction

Envision a long-awaited vacation. You plan and thoroughly schedule a family vacation in another part of the country. Upon arriving at the airport you see several queues for domestic checkin, international checkin, and then of course the extremely long security queue. After you have checked the family in, you proceed to the security queue, knowing all too well how painful the wait in line will be. When you arrive at the security checkpoint, the security monitor quickly recognizes your premier status and escorts you quickly to the priority queue, passing up the delay of the general security queues. Here you are passed directly through security, and you hurry the family quickly off to the gate to enjoy your vacation.

Similar to the example above, quality of service (QoS) on the Cisco Catalyst 4500 Supervisor Engine 6-E (Supervisor Engine 6-E) is a tool that is used to provide preferential treatment to specific traffic as it passes through the switch. Over time and with the advancements in hardware and software technology, a number of QoS tools have become available. QoS in itself is not one feature, but a collection of features that when combined provide a powerful way to identify different classes of traffic, prioritize them, and then schedule the traffic based on this prioritization.

This document will provide a high-level overview of QoS capabilities of the Supervisor Engine 6-E which is part of the Cisco Catalyst 4500E family. It will explain what the features are and how they work.

### Where Is QoS Performed?

The Cisco Catalyst 4500E performs all QoS on the supervisor engine. This enables the Cisco Catalyst 4500E line cards to expand their QoS feature set and capabilities by simply upgrading one piece of hardware, the supervisor engine. This is investment protection at its best, enabling all line cards purchased as far back as 1999 to perform enhanced QoS capabilities built into the latest supervisor engine, the Supervisor Engine 6-E. This flexibility is a function of the centralized architecture provided by the Cisco Catalyst 4500E, the most widely deployed modular switch to date.

### QoS Enabled by Centralized Architecture

The Supervisor Engine 6-E uses centralized application-specific integrated circuits (ASICs) that provide high-performance, scalable service capacity and superior investment protection. This technology provides advanced QoS capabilities in the Supervisor Engine 6-E, which in turn extends these capabilities to all line cards.

### Traffic Prioritization Overview

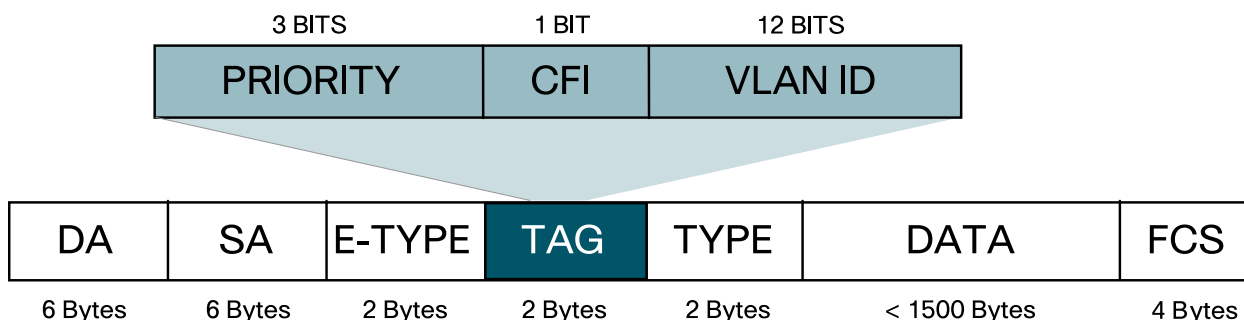
When data is sent through a network, it can be tagged with a priority value. When the data passes through a network device, the network device uses that priority value to determine how it should treat the packet. Data can be tagged with a priority value as described in the following sections.

### Class of Service

When a packet is transmitted out an Ethernet port, it has an Ethernet header attached to it. This Ethernet header can include an optional VLAN tag (also referred to as an IEEE 802.1Q VLAN tag). Within the VLAN tag is a 3-bit field

called the class-of-service (CoS) field. These 3 bits can be manipulated to yield eight different priority values. Figure 1 shows where in the Ethernet header the priority bits are found.

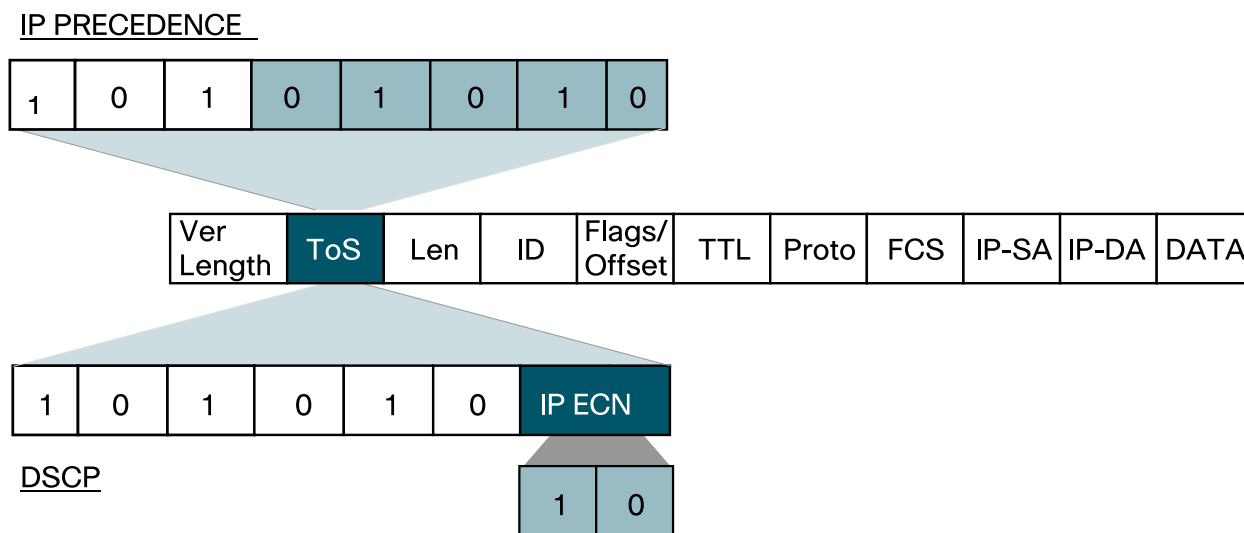
**Figure 1.** 802.1q tag in Ethernet Header



**Type of Service**

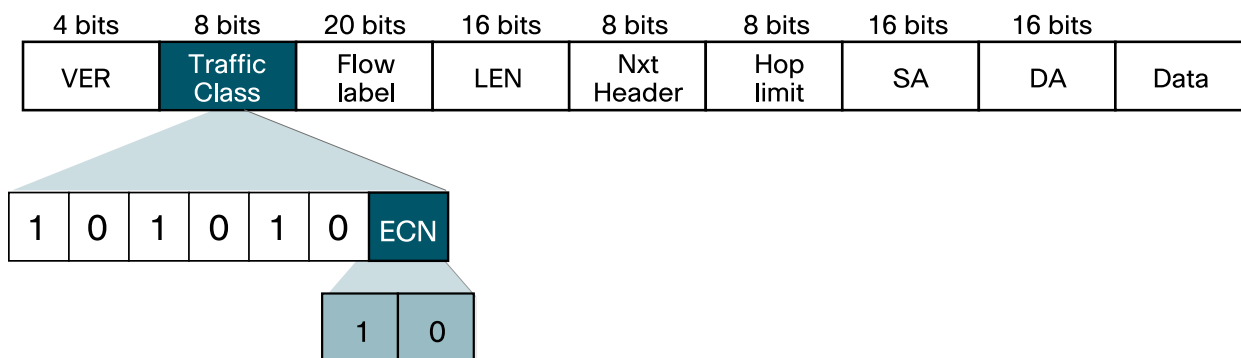
Built into every IP packet is an IP header, and like in the Ethernet example earlier, the IP header also contains a field that defines a priority value for this packet. This field is the type-of-service (ToS) field, an 8-bit field. There are two ways to set a priority value in the ToS field. One method, called IP precedence, uses the 3 most significant bits of the ToS field to yield eight priority values. Differentiated services code point (DSCP) is a second method for assigning a priority to an IP packet. DSCP uses the 6 most significant bits of the ToS field to yield 64 different priority values. The Cisco Catalyst 4500E supports DSCP using the DiffServ model RFC 2475. Figure 2 shows where the ToS bits are found in the IP header.

**Figure 2.** DSCP and IP Precedence in IP Header



**IPv6 Traffic Class**

The IPv6 header also contains a QoS marking field termed traffic class. This field can be manipulated to mark or retain QoS markings of an IPv6 packet.

**Figure 3.** Traffic Class in IPv6 Header

### Categories of QoS Features

Explaining what features are available on the Supervisor Engine 6-E is best served when categorizing them into one of the following groups:

- Classification
- Policing
- Marking
- Queuing
- Congestion avoidance
- Scheduling

The use of **classification** provides a way for the switch to identify specific traffic so that it can determine what level of service needs to be applied to that data. Identification can be achieved by a number of means, such as inspecting primary fields in the packet header or looking at the port of arrival. The main classification tool provided by the Supervisor Engine 6-E is access control lists (ACLs).

The act of **policing** in the switch provides a means to limit the amount of bandwidth that traffic traveling through a given port, VLAN, or collection of VLANs on a port can use. Policing works by defining an amount of data that the switch is willing to send or receive. The policing policy uses an ACL to identify the traffic to which the policer will be applied. Multiple policing policies can be active in the switch at any one time, allowing an administrator to set different rates for different classes of traffic. Policing can be set up so that it rate limits all traffic entering a given port, VLAN, port and VLAN pair, or flow to a given rate.

**Marking** is the action of changing the priority setting of the packet. Each packet consists of data and a header. The header contains, among other things, information such as where the data has come from (the sending device's source address) and where the data is destined (the target device's destination address). Built into the header is the priority value that can be used to indicate to switches and routers in the network path, the priority of that piece of data. The Supervisor Engine 6-E has the ability to change that priority value (increase or decrease it) if required based on any policies that the network administrator might set.

**Queuing** (also known as congestion management) provides a way to temporarily store data when the received rate of data is larger than what can be sent. Like in the earlier airport example, the supervisor engine will use an egress queue to place data into a temporary holding area until the data is scheduled to be forwarded. Memory is allocated to each queue, which provides the buffer space for data awaiting service. One primary advantage of the Cisco Catalyst 4500E CenterFlex architecture is that the number of queues and the amount of buffering available (per port per queue) are solely dependent upon the supervisor and not determined by the line card in use. Within the queue packets are not referenced based on byte count, but rather by number of packets.

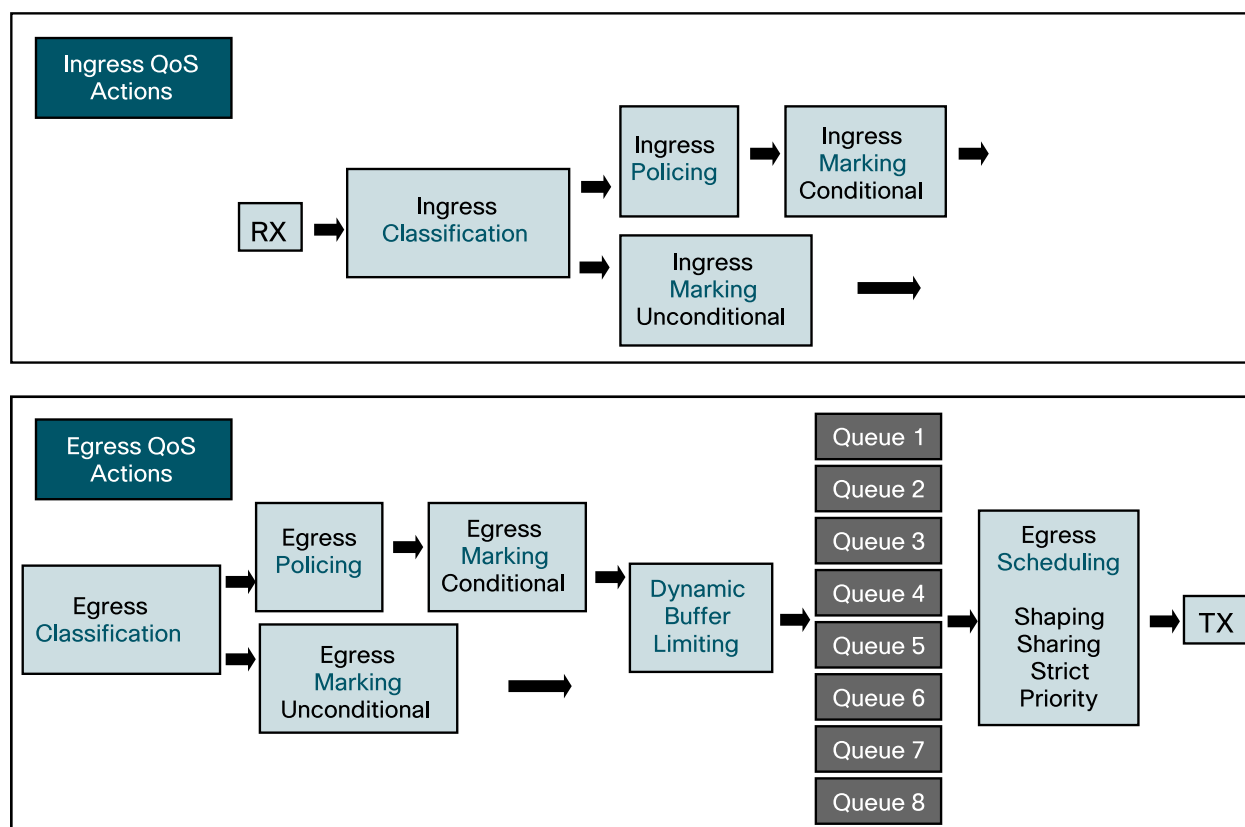
Managing the queues and buffers is the primary goal of **congestion avoidance**. As a queue starts to fill up with data, it is important to try to make sure that the available memory in the queue does not fill up completely. If this happens, subsequent packets coming into the port will simply be dropped, irrespective of the priority that they could have received. This could have a detrimental effect on the performance of critical applications. For this reason, congestion avoidance techniques are used to reduce the risk of a queue from filling up completely. Queue thresholds are used to trigger when certain levels of occupancy are met. Multiple congestion avoidance techniques are implemented in the Cisco Catalyst 4500E after a threshold has been crossed. In addition to standard Waited Tail Drop, the Cisco Catalyst 4500E features an advanced congestion avoidance algorithm, Dynamic Buffer Limiting (DBL). When a threshold has been crossed, the system will isolate nonadaptive or belligerent flows, which consume the buffers, and drop lower priority data while trying to keep as much of the higher priority data resident in the queue. The congestion avoidance technologies used on the Supervisor Engine 6-E include DBL and tail drop.

**Scheduling** is the QoS mechanism used to empty the queues of data and send the data onward to its destination. The scheduling options available in the Supervisor Engine 6-E are shaping, sharing, and strict priority queuing.

To compare QoS on previous Cisco Catalyst 4500E supervisor engines, visit [http://www.in.cisco.com/esg/gsbu/c4000/docs/catalyst4500E\\_QoS\\_overview\\_internal.doc](http://www.in.cisco.com/esg/gsbu/c4000/docs/catalyst4500E_QoS_overview_internal.doc).

Now that the individual groups of QoS have been pointed out, it is important to note the process in which these actions are carried out.

**Figure 4.** QoS Flow in Supervisor6-E



### The QoS Feature Toolkit

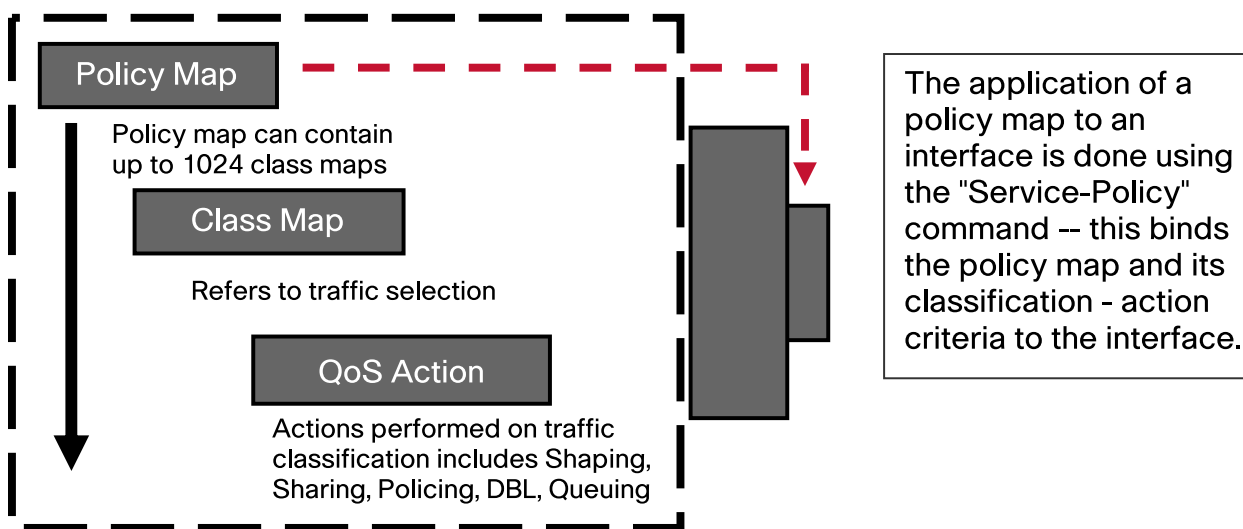
The Supervisor Engine 6-E has been primed with an extended list of enhanced QoS functions far more advanced than that of its predecessors. This feature-rich toolkit puts you in control to better manage and prioritize traffic. The

following sections will explain in detail the enhanced features that make up the Supervisor Engine 6-E's QoS feature toolkit.

### Modular QoS Command-Line Interface

The Supervisor Engine 6-E has been simplified to provide the modular QoS CLI (MQC) command structure that is also found in Cisco IOS Software running on Cisco routers. The normal rules of configuration are such that a class map is built incorporating the ACLs that identify the traffic that will have QoS applied to them. The class map is then referenced within a policy map, which contains the QoS policy that will be applied to the switch port (or VLAN). The policy is then applied to the physical or logical interface. A high-level view of this process is shown.

**Figure 5.** Modular QoS CLI



There are very few differentiators between what MQC CLI routers and the Supervisor Engine 6-E perform. Some of the more noticeable differences between previous supervisor engines and the Supervisor Engine 6-E are:

- Trust
- Internal DSCP
- Table maps
- Sequential vs. parallel classification
- Priority queue placement

Previously supervisor engines relied on "trust" to classify traffic; however, this does not fall into the MQC CLI construct. MQC provides a more flexible capability, table maps. Another difference is the "internal DSCP" value used within the switch to place packets in the proper queue. The Supervisor Engine 6-E does not use "internal DSCP"; rather, it relies on explicit matching of QoS values using class maps so that packets can be placed in the correct queue. Also, note that there is no specific priority queue: it is not queue 3 or queue 1. The priority queue is simply configured within a class; therefore, it is not tied to a specific queue. One final difference is that of classification. The Supervisor Engine 6-E provides sequential classification rather than parallel. This allows the network administrator to classify traffic at egress based on the ingress markings. Based on these changes, QoS CLI will now be more contiguous on the Supervisor Engine 6-E as it will now have standard Cisco MQC CLI, making configuration management that much simpler.

### Ingress QoS: Default Actions

First and foremost, QoS does not need to be enabled on the Supervisor Engine 6-E, it is on by default in compliance with the MQC construct.

When a packet arrives at an interface, there are two options to take into consideration: is there a policy attached or not? If the packet arrives with or without a marking and there **is not** a policy attached to the interface, packets will flow through the switch untouched. There are no questions as to where the packet came from or if it has a valid marking. If the packet arrives with or without a marking, and a policy **is** attached to the interface, the packet will only then be subject to the policy classification.

### Ingress QoS: Table Map

In classic supervisor engine, there is a global table to map the CoS to DSCP and DSCP to CoS, but with the introduction of table maps, we can now have a per-policy or per-interface configuration where CoS can be independently mapped to DSCP values, and we do not have to rely on a single global table. The default mapping is still available if a user does not define a table map. Here are some examples of where a table map can be used.

This is how a table map is defined:

```
C4500E(config) # table-map table-map1
C4500E(config-tablemap)#?
table-map configuration commands:
  default the default behavior for setting value not found in the table map
  exit Exit from table-map configuration mode
  map map to_value from from_value
  no Negate or set default values of a command
C4500E(config-tablemap)# map from 40 to 5
C4500E(config-tablemap)# map from 24 to 3
```

### Ingress QoS: Trusted Boundary

Previously, the trust boundary feature dynamically modified the port trust state to trust the packet marking (CoS, DSCP) only if the presence of a Cisco IP phone is detected on that port using Cisco Discovery Protocol. The trust state was then used to modify the queuing and scheduling of packets.

The Supervisor Engine 6-E supports MQC, which does not support trust, as mentioned above. With this said, trusted boundary has been implemented on this platform to the same specifications as previous supervisor engine. If the trusted boundary feature determines that a Cisco IP phone is not detected on a port (and hence the packet markings received on the port should not be trusted), then the packet marking values (DSCP, CoS) are treated as the port default values instead of the actual values in the packet. This will place all the traffic on the port in the "class default" queue. If, however, a Cisco IP phone is detected, the markings and policy attached to the port will be used to provide the correct packet treatment (that is, priority queue for voice traffic).

For example, on the port on which trusted boundary is configured, if there is a policy attached as given below:

```
int gi 4/1
  QoS trust device cisco-phone
  service-policy input trusted_boundary

class-map dscp24
  match dscp 24

class-map dscp46
  match dscp 46

policy-map trusted_boundary
  class dscp24
    set dscp 40
```

```

class dscp46
  set dscp 63

class class-default

```

If a packet with DSCP 24 is received:

- When a Cisco phone is not discovered, the packet matches "class default."
- When a Cisco phone is discovered, the packet matches class map "dscp24."

### Ingress and Egress QoS: Classification Parameters

Classification as defined above, provides a way for the switch to identify specific traffic so that it can determine what level of service needs to be applied to that data. Below are the criteria that the Supervisor Engine 6-E can match or classify traffic upon. These classifications far exceed in number and scope the previous Cisco Catalyst 4500E supervisor engines. (See Table 1.)

**Table 1.** Classification Parameters

L2 Classification ARP/RARP for IPv4	IEEE 802.1Q CoS	QoS Group					
<b>L2 Classification for NONIP protocols</b>	IEEE 802.1Q CoS	QoS Group	SMAC	DMAC	Ethertype (non- IP)	L2 for IPv4/IPv6	ToS/Traffic Class IPv4/IPv6
<b>IP Classification</b>	IEEE 802.1Q CoS	QoS Group	IP Source	IP Dest	IP DSCP/Traffic Class	IP Protocol IPv4/IPv6	TCP/UDP source/dest port
	ICMP Type/Code	IGMP Type	IP (non-initial) Fragment	Tiny Fragments	TCP Flags		

**Note:** It is not possible to classify a packet based on both L2 and L3 data in the packet (for example, MAC DA and IP address together). However, it is possible to classify based on the following:

- Match on L2 CoS and IP information
- Match on L2 information and IP ToS Match on MAC source and IP information

In addition to the added classification types, the Supervisor Engine 6-E provides enhanced classification. Prior to the Supervisor Engine 6-E, input and output QoS classification happened in parallel; therefore, any marking modified by the input policy could not be used in output QoS classification. With the Supervisor Engine 6-E, the output QoS classification happens after input QoS processing has taken place. This provides separation between ingress and egress classification, which provides the administrator the ability to classify based on ingress markings.

Here are some CLI examples of what parameters can be classified on the Supervisor Engine 6-E.

```

C4500E(config-cmap) #match

access-group  Access group
any           Any packets
cos          IEEE 802.1Q/ISL class of service/user priority values
dscp         Match DSCP in IP(v4) and IPv6 packets
not          Negate this match result //Not Supported
precedence   Match Precedence in IP(v4) and IPv6 packets
protocol     Protocol
qos-group    Qos-group //Only supported for egress classification

```

```

C4500E(config-cmap) #match protocol ?

arp
ipv4
ipv6

```

### Ingress and Egress QoS: 802.3x Flow Control and Thresholds

Oversubscribed ports on the Cisco Catalyst 4500E use 802.3x flow control, more commonly known as pause frames, to control congestion of stub ASICs. Front panel ports connecting to stub ASICs on the line card can be up to 8:1 oversubscribed. The stub ASICs can send and receive pause frames, which cause the device receiving the pause frame to halt all traffic for approximately 33 microseconds. This provides enough time to clear the miniscule buffer on the line card and begin to forward traffic regularly. With 8:1 oversubscribed line cards, this can translate to a minimum bandwidth of 125 Mbps per port, given that all ports are sending traffic at line rate and at the same packet size.

If a port is paused, then all the queues of that port pause too, and no packets will be transmitted. Once the port gets unpaused, it can send a large burst. To limit the amount of bursting, a configurable burst parameter is used in shaping (per queue for queue shaping and per port for port shaping). Burst configuration controls how much over its shape rate a port/queue can transmit.

### Ingress and Egress QoS: Marking

Marking is the process of setting values in the QoS fields of a packet. This includes ToS/traffic class byte in the IPv4/IPv6 header (these fields are interpreted as {DSCP, ECN}), the CoS field in the IEEE 802.1Q header. When a marked packet traverses the network, the packet marking provides the QoS treatment.

The Supervisor Engine 6-E supports marking the following bits in packet headers:

- IEEE 802.1p/802.1Q CoS
- IPv4 ToS/DSCP
- IPv6 traffic class/DSCP
- IP ECN
- IEEE 802.1p/802.1Q CFI/DEI bit (future software releases)

The Supervisor Engine 6-E supports three different types of marking; Unconditional marking with a value, unconditional marking using a table map, conditional marking using a policer.

### Unconditional Marking with a Value

The Supervisor Engine 6-E supports marking all of the QoS related markings in the packet mentioned above unconditionally with a specific value. The Supervisor Engine 6-E can mark all traffic on an interface or VLAN

unconditionally to a specified value. Here are some examples of unconditional marking on the Supervisor Engine 6-E. Also, multiple-field marking is supported. What that means is both CoS and DSCP can be modified simultaneously using a policy on the Supervisor Engine 6-E.

```
C4500E(config-pmap-c)# set ?
cos          Set IEEE 802.1Q/ISL class of service/user priority
dscp        Set DSCP in IP(v4) and IPv6 packets
precedence   Set precedence in IP(v4) and IPv6 packets
qos-group    Set QoS Group
```

```
C4500E(config-pmap-c)# set cos 2
C4500E(config-pmap-c)# set dscp 46
```



#### Unconditional Marking Using a Table Map

The Supervisor Engine 6-E supports marking all of the QoS related markings in the packet mentioned above unconditionally with a value derived from the same field or any another field in the packet. This is accomplished using a table map. This table map is indexed using any of the QoS markings or a QoS group, and the result of the table map lookup is used to mark any of the applicable QoS markings.

For example, the DSCP field in the packet can be derived from the incoming DSCP/traffic class, incoming CoS, QoS group (in output only), IP ECN, and so on.

The number of table maps supported depends on the usage of the policies and the table maps. By default the Supervisor Engine 6-E supports 512 entries for each marking table (DSCP/CoS) per direction for all table maps. For example, the Supervisor Engine 6-E supports:

- 64 different table maps with each one mapping the 8 **CoS** values to DSCP (64 \* 8)

AND

- 8 different table maps with each one mapping the 64 **DSCP** values to another DSCP ( 8 \* 64)

and so on.

Here are some examples of unconditional marking using table maps.

#### Unconditional table-map based marking

```
C4500E(config-pmap-c)# set cos dscp ?
    table Set fram cos value from L3 dscp based on table map

C4500E(config-pmap-c)# set cos dscp table table 1

C4500E# show table-map table1
    Table map table 1
    from 24 to 3
    from 40 to 5
    from 46 to 5
    from 48 to 6
    default copy
```



#### Unconditional table-map based marking

```
C4500E(config-pmap-c)# set dscp cos ?
    table Set Packet ip dscp from cos based on table map

C4500E(config-pmap-c)# set dscp cos table table2

C4500E# show table-map table2
    Table map table2
    from 1 to 8
    from 3 to 24
    from 5 to 46
    from 6 to 48
    default copy
```



#### Conditional Marking Using Policing Result

The Supervisor Engine 6-E supports marking all of the QoS related markings in the packet mentioned above based on result of policing action: that is, whether the packet is conforming, exceeding, or violating the policing rate and

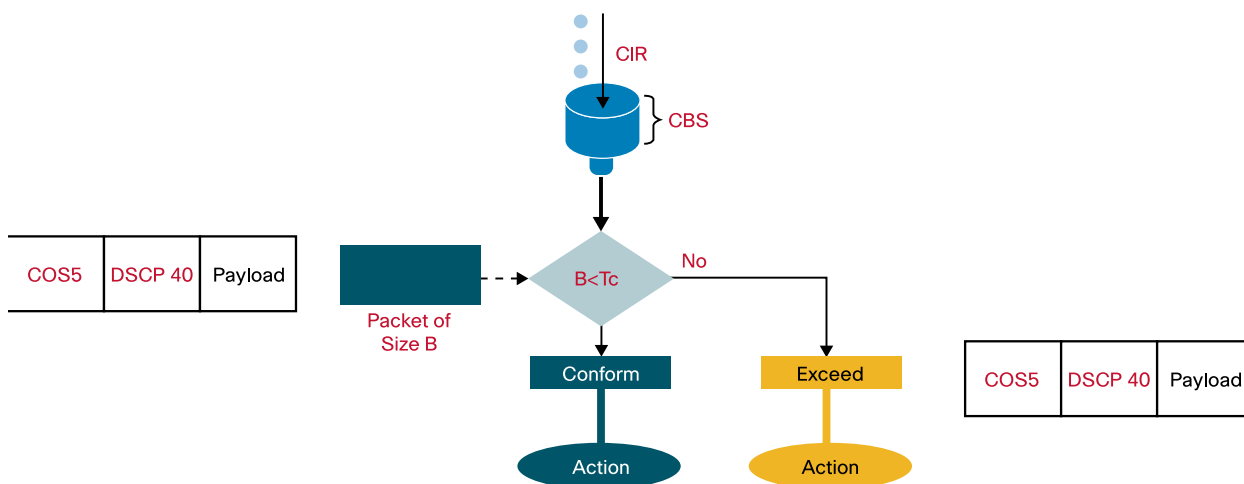
burst. In addition, the conditional marking via policing result is either a specific value or derived from other QoS markings via table mapping. This is also a function of the Supervisor Engine 6-E that is not available on previous supervisor engines. The Supervisor Engine 6-E's ability to mark down not only DSCP but also CoS and other values within a policer provides more flexible QoS capabilities.

Here is an example of conditional policer based marking.

### Conditional policer result based marking

```
C4500E(config-pmap-c)# police cir 32000 conform-action transmit
                        exceed-action set-cos-transmit 0 set-dscp-transmit 0
```

```
drop                drop packet
set-cos-transmit    rewrite packet cos and send it
set-dscp-transmit   set dscp and send it
set-prec-transmit   rewrite packet precedence and send it
transmit            transmit packet
```



### Ingress and Egress QoS: Policing

Policing enforces a maximum transmission rate in order to conform to a contract or service level agreement (SLA). Policing can drop or mark based on the QoS values in the traffic (DSCP, CoS, and so on). This provides lower priority to traffic that is not conforming to the SLA.

Policing can be applied to a group of flows that are transiting a port and/or VLAN; this is called "aggregate policing." An aggregate policer meters all packets on a given port and/or VLAN. An aggregate policer can be "per-interface" or "named" aggregate policer. Per-interface policers are distinct for each port and/or VLAN with which they are associated (via the QoS policy). Even if the same policy is used, each association of the policy instantiates a new policer associated with the port and/or VLAN. A named policer is shared among all ports and/or VLANs with which it is associated.

The Supervisor Engine 6-E supports the following policer types for the named and individual policers:

- Single rate policer two color marker
  - This type of policer is configured with a committed rate (CIR) and normal burst. It has two actions, conform and exceed, where packets can be either green or red. A packet is marked green if it doesn't exceed the CIR, and red otherwise.
  - This is the only form supported in the previous supervisor engines of the Cisco Catalyst 4500E family.
- Single rate three color marker (srTCM) (RFC 2697)

This type of policer is configured with CIR, committed burst size (Bc), and excess burst size (Be), to be either green, yellow, or red. A packet is marked green if it doesn't exceed the Bc, yellow if it does exceed the Bc but not the Be, and red otherwise.

This type of policer supports both color blind mode and color aware mode. This just insinuates that the packet is inspected for markings prior to being marked down by the policer: that is, if the packet is red and exceeds Be, it will be marked red, or if the packet was premarked as red, it will be left alone.

- Two rate three color marker (trTCM) (RFC 2698)

Commonly known as a 2R3C policer, this type of policer supports a committed rate (CIR) and a peak rate (PIR) and has conform, exceed, and violate actions. A packet is marked green if it does not exceed the CIR and yellow if it does. The packet is marked red if it exceeds the peak information rate (PIR). This policer type also supports the color blind mode and color aware mode. This type of policer also allows for multiple marking: that is, marking DSCP and CoS for violating traffic.

- Packet rate policer

Rather than count bytes, this policer type will count based on packet number; this is very useful for policing CPU bound traffic.

**Note:** Named aggregate policers, microflow policing, packet rate policers, and color aware mode, are not supported in the first release of the Supervisor Engine 6-E software.

### Policer Types

The policing algorithm used on the previous supervisor engines is the same as the one used on the Supervisor Engine 6-E.

The Supervisor Engine 6-E supports 16384 (16 x 1024, 16K) single rate, single burst policers. The 16K policers are organized as 8 banks of 2K policers. The policer banks are dynamically assigned to be input policer bank or output policer bank by the software depending on the QoS configuration.

That is, the 16K policers are dynamically partitioned by software, as shown in Table 2.

**Table 2.** Input / Output policers on Supervisor6-E

Input Policers	Output Policers
0	16K
2 K	14 K
4 K	12 K
6 K	10 K
8 K	8 K
10 K	6 K
12 K	4 K
14 K	2 K
16 K	0

**Note:** The numbers in Table 2 represent individual policer entries in the hardware that support a single rate and single burst parameter. Based on this, the Supervisor Engine 6-E supports the following number of policers:

- 16K single rate policer with single burst (two color marker)
- 8K single rate three color marker (srTCM)
- 8K two rate three color marker (trTCM)

The accuracy of the policing algorithm is approximately plus/minus 0.75 percent on either side of the configured policing rate.

### Policing Actions

The Supervisor Engine 6-E supports the following policing actions for conforming, exceeding, and violating results from the policer:

- Transmit
- Drop
- Unconditional marking of QoS fields in the packet with a value

The Supervisor Engine 6-E software supports marking packet fields as a result of specific policing result. That is, marking packets only when the packet conforms to the policing SLA or mark down packets only when the packet exceeds the policing SLA. This is an added capability over and above the current supervisor engines. In the previous supervisor engines the packet was unconditionally marked as a separate action before the policing action for a given traffic class.

The Supervisor Engine 6-E also supports multiaction policers wherein multiple fields in the packet are marked as a result of a specific policing result.

### What Value Does the Policer Consider When Policing?

In previous supervisor engines, with IP packets, only the L3 length (as obtained using the IPv4/IPv6 header) is used in the policing algorithm. For all other packets, the L2 length is used in the policing algorithm.

The Supervisor Engine 6-E supports a systemwide option to account for the entire L2 packet length (including, optionally, the IPG). This is a little different from that of previous supervisor engines, where the option is to add a (programmable) fixed L2 length to the L3 length.

In the Supervisor Engine 6-E, the default length is to use the L2 length, including VLAN overhead, for policing without the IPG. This is standard with all true MQC implementations.

The recommended burst value to be used with policers in the Supervisor Engine 6-E is 0.05 to 0.2 of CIR.

Here are some examples of how to configure policers on the Supervisor Engine 6-E.

#### 2 Rate 3 Color policers

```
C4500E(config-cmap) #police cir percent 10 pir percent
50 conform-action transmit exceed-action set-dscp-
transmit 0 violate-action drop
```

#### 1 Rate 2 Color policers

```
C4500E(config-cmap) #police cir percent 10
conform-action transmit exceed-action drop
```

### Egress QoS: Congestion Avoidance

DBL is a congestion avoidance technique used to drop packets before the congestion occurs. Weighted Random Early Discard (WRED) randomly discards packets at specified queue thresholds. Lower priority ToS/DSCP packets are dropped before higher priority packets, hence the weighting in WRED. This action reduces the average queue size and thus allows the switch or router to detect congestion before the queue overflows. WRED is packet based, not flow based like DBL. Because DBL is flow based and not random, it does not affect “well-behaved” flows that are not causing the queues to congest. DBL uses logical flow tables per port/per queue on each interface A DBL flow is

composed of a source/destination IP address, Layer 4 TCP/UDP ports, and a VLAN. During congestion, the logic in the hardware for dropping a packet is based on the flow. DBL is particularly effective with nonadaptive flows (NAFs). An NAF is any flow that does not reduce the traffic rate in response to packet drops. Usually NAFs use UDP connectionless protocol. Some examples include UDP music or video flows, Internet streaming multimedia, and multicast traffic.

DBL is similar to the Flow Based WRED (FRED), which is used on Cisco IOS Software routers, except it is implemented in hardware at full line rate on all supervisor engines. The hardware implementation is important when deployed in Gigabit Ethernet networks versus typical WANs, where maximum bandwidth is usually T-1 speed (1.5 Mbps), DS-3 (45 Mbps), or perhaps OC-4 (155 Mbps).

DBL is supported on all the ports of a Cisco Catalyst 4500E Series Switch. The DBL transmit queue logic is on all transmit queues on every Fast Ethernet, Gigabit Ethernet, or 10 Gigabit Ethernet port. In addition, the protocols on the port are transparent to DBL, which means that you can have routed, switched, access, or trunk ports with Cisco EtherChannel technology or any other protocol configured on them. A Cisco Catalyst 4500E Series Switch with DBL can be used with a Cisco Catalyst 6500 Series Switch that supports WRED. The DiffServ Internet architecture is structured around per-hop behavior (PHB). For example, with a Cisco Catalyst 4500E Series Switch in the wiring closet with uplinks to a Cisco Catalyst 6500 Series Switch in the distribution/core network, the nonadaptive flows (NAFs) have already been controlled using DBL prior to reaching the Cisco Catalyst 6500 Switch. The Cisco Catalyst 6500 Series could then use WRED on those flows that will respond to packet drops.

Referring to the QoS flow shown at the beginning of this paper, DBL acts on a packet flow before the packet is enqueued, avoiding tail drops. The DBL function also occurs after the policing function. DBL and policing are not alternatives to each other. Policing is used to control selected traffic flows by rate limiting them. It is still possible to have Tx queue congestion particularly when bursting occurs. DBL is a QoS congestion avoidance technique specifically designed to prevent Tx queue congestion.

Here is how DBL is configured within a policy:

```
C4500E(config)# policy-map pl
C4500E(config-pmap)# class cl
C4500E(config-pmap-c)# shape average 256000
C4500E(config-pmap-c)# db1

C4500E(config)# interface gigabitethernet1/1
C4500E(config)# service-policy output pl
```

### Egress QoS: Dynamically Allocated Queuing

The Supervisor Engine 6-E allows user configurable queue depth, unlike previous Cisco Catalyst 4500E supervisor engines. Queue allocation is configurable via the "queue-limit" command. These queues are configurable in chunks that begin at 16 entries. When a queue is modified above its default allocation, the amount of queue entries added is taken from the reserve queue. This queue holds the remaining amount of unused queue entries.

#### How Is the Queuing Capacity Dynamically Allocated?

Queue capacity is first allocated based on the type of queue (CPU, drop, uplink, standard port); each type is given a set capacity, shown in Table 3. Then, each slot within the chassis will be allocated an equal share of the remaining queue space. Finally, all line cards of the same port density will allocate to each port an equal portion of the queuing capacity.

The Supervisor Engine 6-E defaults, with no policy attached to a port, to a single queue, which will be used by each port to pass all traffic. This queue is allocated the entire amount of queuing capacity. Upon configuring the policy, queues are dynamically generated based on the classes configured within the policy if the classes are given a

queuing requirement (for example, 6 classes equates to 6 queues if shaping or other queuing command is configured). Policing does not constitute queuing. These queues are then dynamically allocated an equal portion of the queuing capacity.

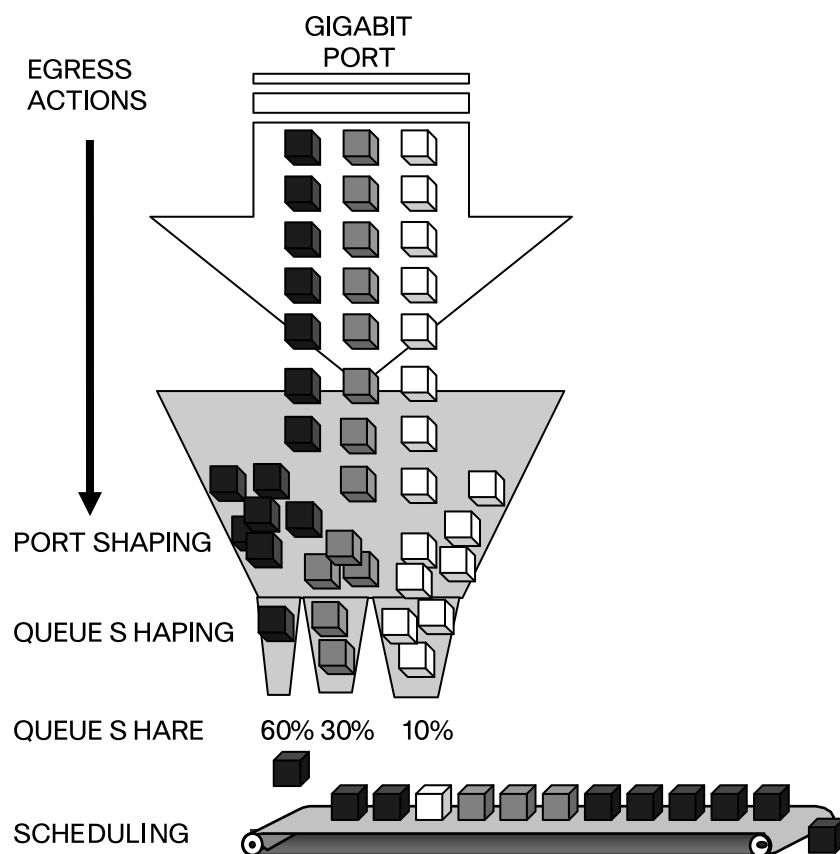
Based on a 10 slot chassis implementing all 8 line-card slots, Table 3 illustrates the number of queue entries available to each line card and each front panel port on the line card. The queue entries shown in Table 3 are equivalent to packets regardless of size.

**Table 3.** Queuing on Supervisor6-E

Queuing Allocation	
<b>Total Queue entries</b>	524288
<b>Drop Queue entries</b>	8192
<b>CPU Queue entries (1k each)</b>	65536
<b>Uplink Queue entries (8k each)</b>	16384
<b>Free Reserve Queue entries</b>	114688
<b>Allocation for all slots</b>	319488
<b>Available per slot</b>	39936
Line Cards	
<b>48 Port Line cards</b>	832 Queue entries per port
<b>*New 6 Port Line card</b>	6656 Queue entries per port

### Egress QoS: Port and Queue Scheduling

Transmit queue scheduling is the process of selecting one of the 8 transmit queues that is eligible to transmit the next packet. If a queue or port is under its shape value, it will then be given its share of bandwidth. Since the Supervisor Engine 6-E supports port and queue scheduling, it is important to note the scheduling hierarchy illustrated below.

**Figure 6.** Egress Queuing

**Note:** Port shaping is not available on the Supervisor Engine 6-E.

Logically, as shown in the above diagram, transmit queue shaping is enforced before sharing. That is, only when a queue is below its shape rate is it considered for scheduling to guarantee its bandwidth sharing rate.

The Supervisor Engine 6-E also supports optional inclusion of VLAN tag, any internal headers, and IPG toward sharing/shaping computation. By default, the L2 encapsulation length is included as part of sharing/shaping, which is identical to previous supervisor engines. This means that the VLAN tag size is included along with the packet length in the sharing/shaping algorithm computation, and IPG is not included. Note that the shape rate configured reflects that the overhead is taken into account.

## Shaping

Transmit queue or port shaping is the act of buffering traffic within a port or queue until it is able to be scheduled. Shaping smoothes traffic, making traffic flows much more predictable. It helps ensure that each transmit queue is limited to a maximum rate of traffic. Shaping is a credible alternative to policing, which drops all traffic exceeding the policer conditions without remorse.

The Supervisor Engine 6-E supports virtual time sharing and shaping (VTSS), the same algorithm used in previous supervisor engines. One of the major differences in shaping from previous supervisor engine is that the shaping algorithm now supports a "shape burst" parameter. This is extremely useful if the link was previously paused. When the link is paused (thus causing traffic on all queues to back up) and then becomes unpaused (thereby making all queues below their shape rate), the burst parameter smoothes traffic on the link by allowing it to catch up simply by bursting traffic.

The shaping accuracy is within 0.75 percent above or below the configured rate.

Here is an example that shows how to configure class-based shaping on the Supervisor Engine 6-E.

```
C4500E#show policy-map p5
Policy Map p5
  Class c1
    shape average percent 32

C4500E(config)#int gigabitethernet 3/3
C4500E(config-if)#service-policy input p5
% A service-policy with queuing actions cannot be attached in input direction.
C4500E(config-if)#service-policy output p5
C4500E(config-if)#end
```

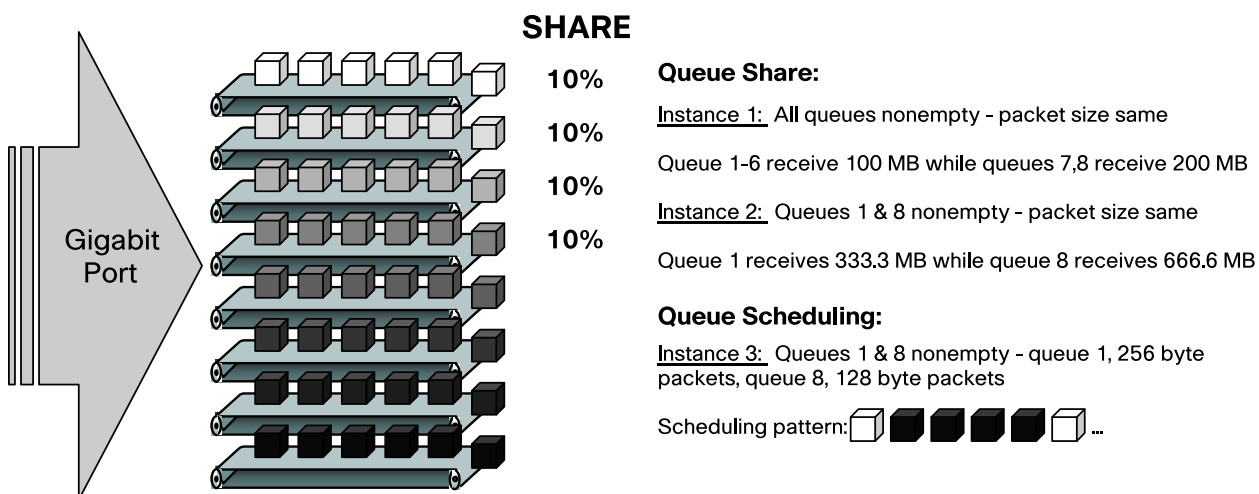
## Sharing

Transmit queue sharing is the process that determines how the output link bandwidth is shared among multiple queues of a given port when the aggregate queue bandwidth is greater than the link bandwidth. This will then provide preferential treatment to one class of traffic at the expense of other classes of traffic. Sharing controls the minimum link bandwidth guaranteed for a given transmit queue.

Smooth round robin is the method used for sharing port or queue bandwidth on the Supervisor Engine 6-E. Sharing as defined on previous supervisor engines is accomplished using round robin. In this case, when using the supervisor engine in conjunction with linecards having oversubscribed stub ports, the bandwidth ratio between 2 ports each having 50 percent share (one with 9k byte packet size the other with 64 byte packet size) has an actual ratio of 1 to 23: that is, almost 23 times more bandwidth is consumed by the jumbo packet port. Different byte counts on different subports resulted in unequal bandwidth among the subports. With the Supervisor Engine 6-E, this is taken into consideration. Weighted round robin scheduling is used to provide each queue with a share of the bandwidth proportional to its configured weight, or by simply specifying the bandwidth percentage. Please reference the following example:

There are 8 queues on a 1 Gbps port. 6 queues are to each receive 10 percent of the link bandwidth (queues 1-6), and 2 queues (queue 7 and 8) are to each receive 20 percent of the link bandwidth. Shown in instance 1, if all queues are active and nonempty all the time, then queues 1-6 each receive 100 Mbps where queues 7 and 8 each receive 200 Mbps. As in instance 2, queue 1 and 8 are the only active queues; then queue 1 receives 333.3 Mbps and queue 8 receives 666.6 Mbps. This ability is termed dynamic bandwidth allocation; at no point will bandwidth go unused.

**Figure 7.** Class Based Weighted Fair Queuing



Here is an example of how to configure sharing within the queues on the Supervisor Engine 6-E.

```
C4500E (config) #policy-map p5
C4500E (config-pmap-c) #class c2
C4500E (config-pmap-c) #bandwidth percent 20
```

### Auto QoS

The Cisco Catalyst 4500E Series Auto QoS feature is used to simplify QoS for VoIP deployments. It is available in Cisco IOS Software Release 12.1(40) SG and later on the Cisco Catalyst 4500E Series supervisor engines. With Auto QoS enabled, ingress traffic is automatically classified and then placed into the appropriate egress queue. Auto QoS should be enabled on ports directly connected to Cisco IP phones as well as uplink ports that will be transmitting VoIP traffic. After being configured, Auto QoS performs the following functions:

- Detects the presence or absence of a Cisco IP phone
- Configures ingress traffic classification
- Enables a service policy that matches all traffic and enables DBL on the interface for congestion avoidance
- Automatically shapes the VoIP traffic

Here is an example of how Auto QoS can be configured on the Supervisor Engine 6-E.

```
C4500(config)interface gig 1/1
C4500(config)auto qos voip ?

cisco-phone
trust
```

Auto QoS is not supported on EtherChannel on the Supervisor Engine 6-E.

### QoS On Etherchannel

On EtherChannel, since queuing cannot be enabled for a port channel interface, we typically need two policies: one policy to do marking and policing if required and the second policy to do any queuing. The policing and marking policy is applied on the port channel interface, and queuing policy is applied on the physical member ports of the port channel. Here is an example of a policy on EtherChannel.

```
policy-map p1
  class c1
    police cir 200m

policy-map p2
  class c1
    priority
  class c2
    bandwidth remaining percent 20
    queue-limit 4096
```

```
interface pol
  service-policy output p1

interface gi 1/1
  channel-group 1
  service-policy output p2

interface gi 1/2
  channel-group 1
  service-policy output p2
```

## For More Information

For more information about Cisco IOS Software Release 12.2(50)SG on the Cisco Catalyst 4500E Series Switch, visit <http://www.cisco.com/en/US/docs/switches/lan/catalyst4500/12.2/50sg/configuration/guide/QoS.html>.

## Summary

QoS is no longer an option in networks today; it has become a requirement due the extensive use of unified communications. The extensive QoS toolkit provided by the Supervisor Engine 6-E provides the ability to prioritize traffic through your network enabling voice, video, and data traffic to flow transparently. This paper has covered features and benefits of the Supervisor Engine 6-E's QoS implementation. If you would like further details, visit the latest configuration guide on [cisco.com](http://www.cisco.com).



**Americas Headquarters**  
Cisco Systems, Inc.  
San Jose, CA

**Asia Pacific Headquarters**  
Cisco Systems (USA) Pte. Ltd.  
Singapore

**Europe Headquarters**  
Cisco Systems International BV  
Amsterdam, The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at [www.cisco.com/go/offices](http://www.cisco.com/go/offices).

CCDE, CCSI, CCENT, Cisco Eos, Cisco HealthPresence, the Cisco logo, Cisco Lumin, Cisco Nexus, Cisco Nurse Connect, Cisco Stackpower, Cisco StadiumVision, Cisco TelePresence, Cisco WebEx, DCE, and Welcome to the Human Network are trademarks; Changing the Way We Work, Live, Play, and Learn and Cisco Store are service marks; and Access Registrar, Aironet, AsyncOS, Bringing the Meeting To You, Catalyst, CCDA, CCDP, CCIE, CCIP, CCNA, CCNP, CCSF, CCVP, Cisco, the Cisco Certified Internetwork Expert logo, Cisco IOS, Cisco Press, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Cisco Unity, Collaboration Without Limitation, EtherFast, EtherSwitch, Event Center, Fast Step, Follow Me Browsing, FormShare, GigaDrive, HomeLink, Internet Quotient, IOS, iPhone, iQuick Study, IronPort, the IronPort logo, LightStream, Linksys, MediaTone, MeetingPlace, MeetingPlace Chime Sound, MGX, Networkers, Networking Academy, Network Registrar, PCNow, PIX, PowerPanels, ProConnect, ScriptShare, SenderBase, SMARTnet, Spectrum Expert, StackWise, The Fastest Way to Increase Your Internet Quotient, TransPath, WebEx, and the WebEx logo are registered trademarks of Cisco Systems, Inc. and/or its affiliates in the United States and certain other countries.

All other trademarks mentioned in this document or website 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. (0903R)