



CHAPTER 22

Hierarchical Scheduling and Queuing

The performance routing engine (PRE3 and PRE4) supports a hierarchical queuing framework (HQF) for scheduling and queuing. This HQF architecture enables service providers to manage their QoS services at three or four layers of hierarchy. The scheduler uses the HQF to allocate excess bandwidth among the subscriber queues and logical interfaces. The scheduler services queues based on the maximum rate and bandwidth-remaining ratio you specify.

This chapter describes hierarchical scheduling and queuing, and includes the following topics:

- [Hierarchical Queuing Framework, page 22-1](#)
- [MQC Hierarchical Queuing with 3-Level Scheduler, page 22-5](#)
- [4-Level Scheduler, page 22-10](#)
- [Related Documentation, page 22-12](#)

Hierarchical Queuing Framework

The hierarchical queuing framework (HQF) defines a QoS architecture for implementing hierarchical packet scheduling and queuing on the PRE3 and PRE4. The HQF enables service providers to manage their QoS at three or four levels of hierarchy. The 3-level HQF scheduler uses the following hierarchy:

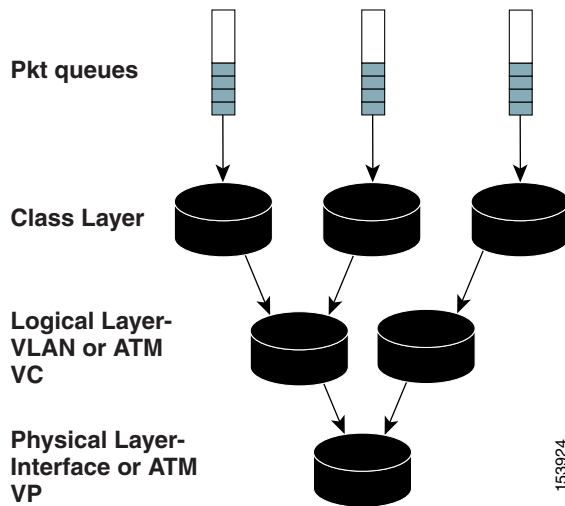
- Physical layer—Used for shaping the physical interface such as the OC-3 port.
- Logical layer—Used to schedule subinterfaces such as a VLAN or PPP sessions.
- Class layer—Used for class queues, defined using the modular QoS command line interface (MQC) policy map.

The 4-level HQF scheduler uses the same hierarchy as above, except that it splits the logical layer into an upper logical layer for sessions and a lower logical layer for subinterfaces. For more information, see the [“4-Level Scheduler” section on page 22-10](#).

The parallel express forwarding (PXF) engine performs all packet-level scheduling using the HQF.

[Figure 22-1](#) shows the 3-level HQF hierarchy.

Figure 22-1 HQF 3 Layers of Hierarchy

**Note**

The PRE1 and PRE2 use the virtual time management system (VTMS) scheduling algorithm and do not support the HQF architecture.

Feature History for Hierarchical Queuing Framework

Cisco IOS Release	Description	Required PRE
Release 12.2(31)SB2	This feature was introduced on the PRE3.	PRE3
Release 12.2(33)SB	This feature was introduced on the PRE4.	PRE3, PRE4

Hierarchical Queuing Framework Scaling

The hierarchical queuing framework (HQF) supports the following interfaces:

- 61,500 logical interfaces
- 16,000 physical interfaces
- Up to 15 queues per interface (2 priority queues [PQs], 12 nondefault queues, and 1 default queue)

QoS Shaping Using HQF

The PRE3 and PRE4 support QoS shaping using the HQF algorithm. The following sections describe how the HQF is used to provide shaping for various QoS models:

- [ATM Virtual Path Shaping Using HQF, page 22-3](#)
- [ATM VC Shaping Using HQF, page 22-3](#)
- [Hierarchical ATM VP and VC Shaping Using HQF, page 22-4](#)
- [Subinterface Shaping Using HQF, page 22-4](#)
- [IP and PPP Session Shaping Using HQF, page 22-5](#)

ATM Virtual Path Shaping Using HQF

A permanent virtual path (PVP) is used to multiplex one or more virtual circuits (VCs). To create a PVP, use the **atm pvp** command in interface configuration mode:

```
atm pvp vpi peak-rate [no-f4-oam]
```

The HQF algorithm treats ATM virtual paths (VPs) as physical interfaces and uses the peak rate you specify to shape bandwidth. The ATM segmentation and reassembly (SAR) mechanism is configured the same as on the PRE2.

The following example shows how to create a PVP with a peak rate of 50,000 kbps:

```
interface atm 7/0/0
  atm pvp 25 50000
```

ATM VC Shaping Using HQF

HQF treats ATM VCs created on the physical interface as logical interfaces and the ATM port as the physical layer. The PRE3 and PRE4 do not support ATM SAR-based VC shaping.

The following examples show how to configure ATM VC shaping for the PRE3 and PRE4. The configuration creates a variable bit rate-nonreal-time (VBR-nrt) VC on the physical ATM interface with a service policy applied to it. HQF shapes the VC traffic according to the sustained cell rate (SCR) of the VC (512 kbps). The service policy applied to the interface creates two separate class queues on the VC: real-time and class-default.

```
policy-map pppoe_vc_out
  class real-time
    police percent 10 200 ms 100 ms conform-action transmit exceed-action drop
    violate-action drop
    priority

  class class-default
    random-detect aggregate
    random-detect precedence values 0 minimum-thresh 10 maximum-thresh 20 mark-prob 10
    random-detect precedence values 1 minimum-thresh 40 maximum-thresh 80 mark-prob 10

interface ATM 1/0/0.1 point-to-point
  pvc 0/110
    vbr-nrt 512 512 94
    encapsulation aal5autopp Virtual-Template1
    service-policy output pppoe_vc_out
```

Hierarchical ATM VP and VC Shaping Using HQF

When VCs are created in a virtual path (VP), HQF treats the VCs as logical interfaces and the VP as the physical interface.

HQF shapes the aggregate traffic from all of the VCs at both the packet and ATM levels (see the “[ATM Virtual Path Shaping Using HQF](#)” section on page 22-3). However, the parallel express forwarding (PXF) engine shapes the VCs at the packet level. Therefore, VC-level ATM shaping is not guaranteed to be compliant to an ATM layer policer at the VC level. Multiple cells from the same VC are sent back at the rate of the VP. The ATM SAR is configured the same as on the PRE2.

The following example shows how to create VBR-nrt VCs in a VP tunnel and apply service policies to each of the VCs. HQF shapes the VCs individually according to their SCR parameter and shapes the aggregate traffic from all of the VCs at the VP rate. The service policies applied to the VCs create class queues on each one of the VCs. Note that unless oversubscription is enabled, the aggregate rates of the VCs cannot exceed the VP rate.

```
interface atm 7/0/0
  atm pvp 25 50000
  pvc 25/100
    vbr-nrt 10000 5000 16
    encapsulation aal5autopp Virtual-Template1
    service-policy output pppoe_vc_out

  pvc 25/101
    vbr-nrt 10000 7000 16
    encapsulation aal5autopp Virtual-Template1
    service-policy output pppoe_vc_out

  pvc 25/110
    vbr-nrt 10000 2000 16
    encapsulation aal5autopp Virtual-Template1
    service-policy output pppoe_vc_out
```

Subinterface Shaping Using HQF

HQF treats subinterfaces (such as VLANs, QinQ, Frame Relay DLCIs, and so on) at the lower logical layer. To shape a subinterface, apply a service policy to the subinterface.



Note

You cannot simultaneously apply service policies to the physical interface and the subinterface.

For example, to shape the aggregate traffic on a VLAN subinterface, apply a hierarchical policy to the subinterface as shown in the following configuration. In this example, the VLAN is shaped at 100 kbps.

```
policy-map child
  class precedence0
    bandwidth percent 10
  class precedence1
    shape average percent 50
    random-detect

policy-map parent
  class class-default
    shape average 100000
    service-policy child
```

IP and PPP Session Shaping Using HQF

HQF treats IP and PPP sessions at the upper logical layer to provide bandwidth sharing and maximum rate shaping. To shape IP or PPP sessions, apply a service policy to the sessions using a virtual template or a RADIUS server.

HQF allows the oversubscription of sessions on a subinterface (such as VLAN, QinQ, or VC) and at the same time also allows oversubscription of the VLAN or VC on a physical port. During congestion, HQF fairly shares the bandwidth first at the subinterface (lower logical layer) and then among the sessions (upper logical layer). HQF takes the bandwidth that was distributed to the subinterface and fairly shares it among the sessions of that subinterface. HQF then takes the bandwidth distributed to a session and fairly shares it among the class queues of that session.

MQC Hierarchical Queuing with 3-Level Scheduler

The MQC Hierarchical Queuing with 3-Level Scheduler feature enables you to configure per-session QoS and subinterface shaping of the aggregate session traffic. This feature provides a flexible packet scheduling and queuing system in which you can specify how excess bandwidth is to be allocated among the subscriber queues and logical interfaces. Rather than allocating an implicit minimum bandwidth guarantee to each queue, the 3-level scheduler uses the bandwidth-remaining ratio parameter to allocate unused bandwidth to each logical queue. The 3-level scheduler services queues based on the following user-configurable parameters:

- Maximum rate—The specified shape rate of the parent queue
- Bandwidth-remaining ratio—The value used to determine the portion of unused, nonguaranteed bandwidth allocated to a logical queue relative to other queues competing for the unused bandwidth



Note

At the class level, the router converts the values you specify for the **bandwidth bps** and **bandwidth remaining percent** commands to a bandwidth-remaining ratio value. The router does not allow you to configure the **bandwidth bps** and **bandwidth remaining percent** commands on the physical and logical layers.

The 3-level scheduler on the PRE3 supports priority propagation by propagating the priority guarantees you configure for subscriber services down to the logical interface level. Therefore, the priority traffic is serviced first at the logical and class level. After servicing the priority traffic bandwidth, the 3-level scheduler allocates unused bandwidth to the logical queues based on the configured bandwidth-remaining ratio. In the default case, the 3-level scheduler allocates an equal share of the unused bandwidth to each logical queue.

For ATM VCs, the 3-level scheduler shares bandwidth proportionally to each VC's bandwidth, if no bandwidth remaining ratio (BRR) or VC weight is configured. For other types of subinterfaces, the scheduler distributes the bandwidth equally, unless BRR is configured. The scheduler uses a default BRR value of 1 if BRR is not specified, except for the ATM logical layer as mentioned above. The logical layer BRR is completely independent from the BRRs configured at the class layer.

The 3-level scheduler supports shaping and scheduling only on the egress interface. The **bandwidth** command must be configured as a percentage of the available bandwidth or as an absolute bandwidth. You cannot concurrently configure the **bandwidth** and **bandwidth remaining** commands on the same class queue or the same policy map.

For more information about the bandwidth-remaining ratio, see the [“Distribution of Remaining Bandwidth Using Ratio”](#) section on page 5-14.

For more information about the 4 level scheduler, see the “4-Level Scheduler” section on page 22-10

Feature History for MQC Hierarchical Queuing with 3-Level Scheduler

Cisco IOS Release	Description	PRE Required
Release 12.2(31)SB2	This feature was introduced and implemented on the Cisco 10000 series router for the PRE3.	PRE3
Release 12.2(33)SB	This feature was introduced on the PRE4 for the Cisco 10000 series router.	PRE4

Prerequisites for MQC Hierarchical Queuing with 3-Level Scheduler

Traffic classes must be configured on the router using the **class-map** command.

Restrictions for MQC Hierarchical Queuing with 3-Level Scheduler

- We recommend that the sum of all priority traffic on a given interface not exceed 90 percent of the physical bandwidth of that interface.
- The 3-level scheduler does not support bandwidth propagation. Therefore, you cannot configure a bandwidth guarantee for any queue other than a priority queue.
- To allow oversubscription provisioning, the admission control check is not performed.
- The sum of all priority traffic running on a given port must be less than or equal to 90 percent of the port bandwidth.

Scheduling Hierarchy

As shown in [Figure 22-1](#), the 3-level scheduler uses the following scheduling hierarchy to allocate bandwidth for subscriber traffic:

- Class layer—The 3-level scheduler uses virtual-time calendars to schedule class queues.
- Logical layer (subinterface, session, or ATM VC)—Virtual-time calendars perform weighted round robin based on the weight of the logical interface and the number of bytes dequeued.
- Physical layer (interface or ATM virtual path)—A real-time calendar ensures that the maximum rate for the class and the logical interface are not exceeded.

By using VP and VC scheduling with existing Cisco 10000 ATM line cards, the scheduler supports priority propagation: cell-based VP shaping in the segmentation and reassembly (SAR) mechanism with frame-based VC scheduling in the performance routing engine 3 (PRE3).

Priority Service and Latency

The 3-level scheduler supports two levels of priority queues; for example, level 1 for voice traffic and level 2 for video traffic.

For a priority class with policing configured, the 3-level scheduler always polices the priority traffic to the rate specified in the **police** command (1000 kbps as shown in the following sample configuration), regardless of whether or not the underlying interface is congested.

```
Router(config-pmap-c)# police 1000
Router(config-pmap-c)# priority
```



Note

The 3-level scheduler does not support the **priority kbps** command.

Latency Requirements

Delay-sensitive traffic incurs a maximum of 10 milliseconds (ms) of latency on edge router interfaces and a maximum of 1 ms of latency on core router interfaces. For interface speeds at T1/E1 and below, the 3-level scheduler services 2 maximum transmission units (MTUs) of nonpriority traffic before servicing a priority packet. Requirements for high-speed interfaces are not as strict as 2 MTUs, but are always bound by 10 ms on edge interfaces and 1 ms on core interfaces.

The 3-level scheduler also supports the minimal latency requirement (2 MTUs of nonpriority traffic in front of priority traffic) at the physical link rate. However, in some cases, it is impossible for the 3-level scheduler to service all competing packets with a latency of 2 MTUs. For example, if many priority packets compete at the same time for bandwidth, the last one serviced may incur latency that is greater than 2 MTUs.

[Table 22-1](#) lists the maximum latency requirements for various interface speeds.

Table 22-1 Maximum Latency Requirements

Interface Speed	Maximum Latency
Greater than 2 Mbps	2 MTU + 6 ms
2 Mbps to 1 Gbps	2 MTU
1 Gbps or greater	1 ms

Priority Propagation with Imposed Burstiness

A single physical interface can have large numbers of logical interfaces and each of these logical interfaces can have both priority and nonpriority traffic competing for the physical link. To minimize latency, the priority traffic of one logical interface has priority over the nonpriority traffic of other logical interfaces, thereby imposing burstiness on the minimum rate traffic of other logical interfaces. The latency that the priority traffic incurs results from the rate constraining the delivered rate of the priority traffic. In many cases, this constraining rate is not the rate of the priority class's parent policy.

For example, suppose a 10 Gigabit Ethernet (GE) interface has 100 VLANs that are shaped to various rates. Each VLAN has a priority class and additional classes configured. Through priority propagation, the scheduler delivers latency to the priority traffic based on the 10 GE rate and not the VLAN rate.

**Note**

The VLAN rate is at most 1 to 2 MTUs of nonpriority traffic in front of priority traffic, which would bound the latency incurred by priority traffic (due to nonpriority traffic) at 1 to 2 MTUs served at the 10 GE rate.

The priority traffic of one logical interface cannot only impose burstiness on other traffic, but also starve other traffic. The only way to prevent the starvation of other traffic is by configuring a policer on the priority queue by limiting the percent of priority traffic to less than 90 percent of the parent bandwidth and the port bandwidth.

Configuring MQC Hierarchical Queuing with 3-Level Scheduler

To configure the 3-level scheduler by allocating excess bandwidth, use the **bandwidth remaining ratio** command. For more information, see the “[Distribution of Remaining Bandwidth Using Ratio](#)” section on page 5-14.

Configuration Examples for MQC Hierarchical Queuing with 3-Level Scheduler

This section provides the following configuration examples:

- [Bandwidth Allocation—Policy Attached to an Interface: Example, page 22-8](#)
- [Tuning the Bandwidth-Remaining Ratio: Example, page 22-9](#)

Bandwidth Allocation—Policy Attached to an Interface: Example

The following sample configuration consists of one policy map named Child with the following traffic classes defined: prec0, prec2, and class-default. The policy is attached to ATM interface 1/0/0.

```
policy-map Child
  class prec0
    bandwidth 300
  class prec2
    bandwidth 100
  class class-default
    bandwidth 50
!
interface atm 1/0/0
  service-policy output Child
```

Assuming that the traffic flow through each class is enough to require maximum possible bandwidth, the 3-level scheduler allocates bandwidth as described in [Table 22-2](#).

Table 22-2 Queuing Presentation—Policy Attached to an Interface

Traffic Class	Bandwidth Ratio	Total Bandwidth Allocated
prec0	6	666 kbps
prec2	2	222 kbps
class-default	1	111 kbps

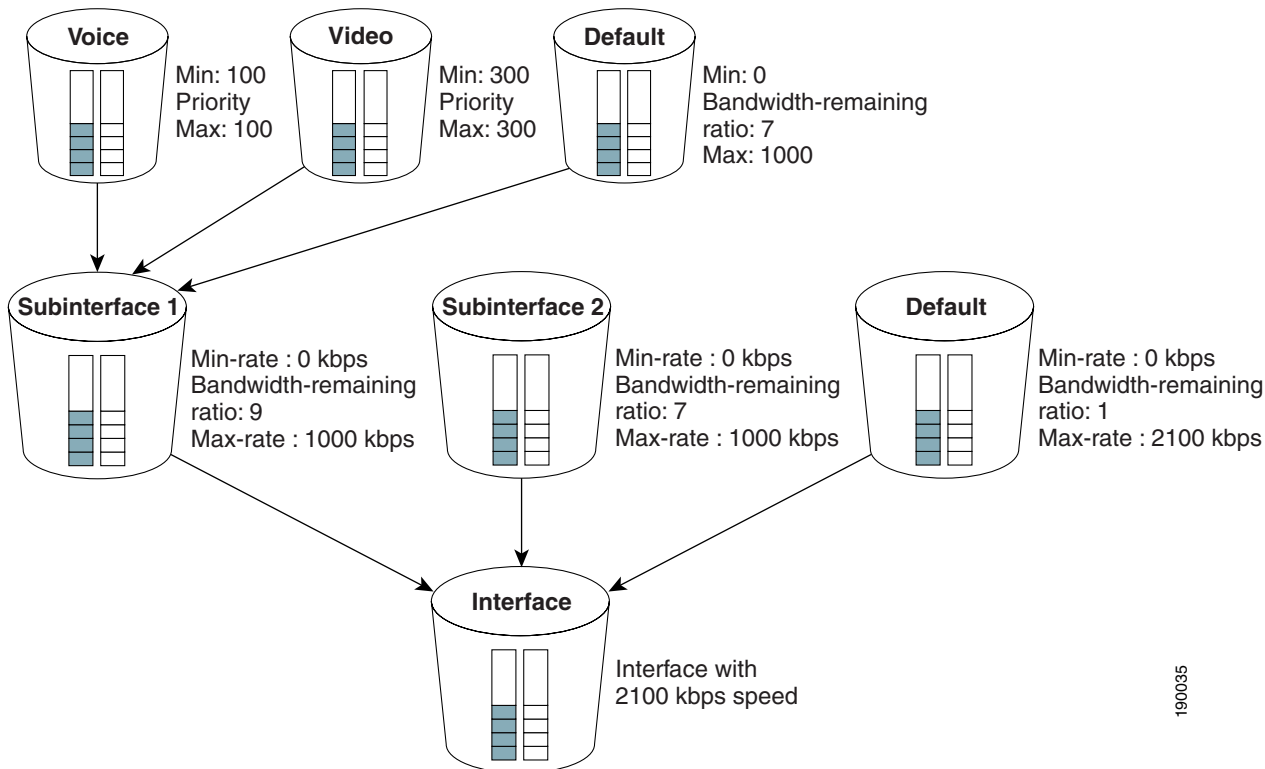
Tuning the Bandwidth-Remaining Ratio: Example

The following sample configuration shows how to tune the bandwidth-remaining ratio using the **bandwidth remaining ratio** command. In the example, the class-default class of Parent1 has a bandwidth-remaining ratio of 9 and the class-default class of Parent2 has a bandwidth-remaining ratio of 7.

```
policy-map Child
  class prec0
    police 100
    priority level 1
  !
  class prec2
    police 300
    priority level 2
  !
policy-map Parent1
  class class-default
    shape average 10000
    bandwidth remaining ratio 9
    service-policy Child
  !
policy-map Parent2
  class class-default
    shape average 1000
    bandwidth remaining ratio 7
    service-policy Child
```

[Figure 22-2](#) shows an example of the queuing presentation based on the above configuration and assuming that the Parent1 policy is enabled on subinterface 1 and the Parent2 policy is enabled on subinterface 2, and that the interface speed is 2100 kbps.

Figure 22-2 Queuing Presentation—Tuning the Bandwidth-Remaining Ratio



Based on the preceding configuration, the 3-level scheduler distributes bandwidth in the following way (assuming that the voice traffic is active on subinterface 1 only and the video traffic is active on subinterface 2 only):

- A total of 400 kbps of bandwidth is used from the interface: 100 kbps-bandwidth guarantee for voice traffic on subinterface 1 and 300-kbps bandwidth guarantee for video traffic on subinterface 2.
- The remaining 1700-kbps bandwidth is distributed across the subinterface-level queues based on their bandwidth-remaining ratios:
 - Subinterface 1 with bandwidth-remaining ratio 9 receives 956 kbps.
 - Subinterface 2 with bandwidth-remaining ratio 7 receives 743 kbps.

4-Level Scheduler

The 4-Level Scheduler feature enables you to configure per-session QoS and subinterface shaping of the aggregate session traffic, just as the 3-level scheduler does. However, unlike the 3-level scheduler, the 4-level scheduler uses the following scheduling hierarchy to allocate bandwidth for subscriber traffic:

- **Class layer**—The 4-level scheduler uses virtual-time calendars to schedule class queues and logical interfaces.
- **Session layer (upper logical)**—Virtual-time calendars perform weighted round robin based on the weight of the logical interface and the number of bytes queued.
- **Subinterface layer (lower logical) (VLAN, QinQ, or ATM VC)**—Virtual-time calendars ensure that the maximum rate for the class and the logical interface are not exceeded.

**Note**

The subinterface layer (lower logical) supports the bandwidth remaining ratio command for Ethernet VLANs and ATM VCs, and the weight command for ATM VCs.

- Physical layer (Ethernet interface or ATM virtual path)—A real-time calendar ensures that the maximum rate for the class and the logical interface are not exceeded.

The 4-level scheduler provides bandwidth sharing and maximum rate shaping among the sessions at the session layer (upper logical) and at the same time among the VLANs and VCs at the subinterface layer (lower logical). The scheduler supports the simultaneous oversubscription of the sessions on a VLAN or VC and of the VLAN or VC on a physical port.

During congestion, the 4-level scheduler does the following:

1. Shares bandwidth fairly at the VLAN, QinQ, or VC level.
2. Shares the distributed VLAN, QinQ, or VC bandwidth fairly among the sessions of that VLAN, QinQ, or VC.
3. Shares the bandwidth distributed to a session fairly among the class queues of that session.

The 4-level scheduler is disabled by default.

**Note**

The router does not convert 3-level queuing hierarchies to 4-level hierarchies. Instead, if 3 levels are needed, then the router uses only 3 levels.

For information about the 3-level scheduler, see the [“MQC Hierarchical Queuing with 3-Level Scheduler” section on page 22-5](#).

Feature History for 4-Level Scheduler

Cisco IOS Release	Description	PRE Required
Release 12.2(33)XNE1	This feature was introduced on the Cisco 10000 series router for the PRE3 and PRE4.	PRE3, PRE4

Related Documentation

This section provides hyperlinks to additional Cisco documentation for the features discussed in this chapter. To display the documentation, click the document title or a section of the document highlighted in blue. When appropriate, paths to applicable sections are listed below the documentation title.

Feature	Related Documentation
Class-based shaping	<i>Cisco IOS Quality of Service Solutions Configuration Guide, Release 12.3</i> Part 4: Policing and Shaping > Configuring class-Based Shaping
Class maps	<i>Cisco IOS Quality of Service Solutions Configuration Guide, Release 12.2</i> Part 8: Modular Quality of Service Command-Line Interface > Configuring the Modular Quality of Service Command-Line Interface > Modular QoS CLI Configuration Task List > Creating a Traffic Class
Policing	Comparing Traffic Shaping and Traffic Policing for Bandwidth Limiting
Policy maps	<i>Cisco IOS Quality of Service Solutions Configuration Guide, Release 12.2</i> Part 8: Modular Quality of Service Command-Line Interface > Configuring the Modular Quality of Service Command-Line Interface > Modular QoS CLI Configuration Task List > Creating a Traffic Policy
QoS service policies	<i>QoS Configuration and Monitoring, Creating Time-of-Day QoS Service Policies</i> Tech Note <i>QoS Configuration and Monitoring, Monitoring Voice over IP Quality of Service</i> Tech Note <i>Site-to-Site MPLS VPN Solution for Service Providers, Service Provider Quality-of-Service Overview</i> Tech Note
Traffic shaping	<i>Cisco IOS Quality of Service Solutions Configuration Guide, Release 12.3</i> Part 4: Policing and Shaping