CHAPTER 6

Configuring QoS on the Layer 3 Queues for the PA-A3 and PA-A6 ATM Port Adapters

In Chapter 2, "Cisco 7200 Series Architecture and Design for ATM Traffic Management," you learned that the PA-A3 and PA-A6 ATM port adapters use both hardware and software queues to support the receive and transmit processing of ATM traffic on a Cisco 7200 series router. The hardware queues, in the form of transmit and receive buffers on the ATM port adapter itself, operate on a first-in first-out (FIFO) basis and are not configurable. The hardware queues operate with the private interface pools (and their receive rings and transmit rings) on the network processing engine (NPE) or network services engine (NSE) for the storage of content during receive and transmit processing.

The software queues operate at Layer 3 and activate when congestion builds on the router and outbound traffic cannot be processed on the transmit ring. Control of the packets awaiting transmit processing passes to the Layer 3 queues according to the quality of service (QoS) policies that you configure. If you do not configure any service policies, then the default behaviors apply. Process-switched packets automatically enqueue to the Layer 3 queues regardless of the state of congestion on the router.

The PA-A3 and PA-A6 ATM port adapters support per-VC queueing at Layer 3. With this design, you can prevent any single VC from starving other VCs for resources.

This chapter provides a brief introduction and some guidelines for configuring the IP to ATM class of service (CoS) features on the PA-A3 and PA-A6 ATM port adapters. The chapter does not describe the full details about the supported QoS implementations in the Cisco IOS software, or all of the possible configuration options for QoS on the PA-A3 and PA-A6 ATM port adapters.

For in-depth information about configuring QoS, refer to the *Cisco IOS Quality of Service Solutions Command Reference* and the *Cisco IOS Quality of Service Solutions Configuration Guide* publications for your software release. Refer to the "Related Documentation" section on page 6-14 for additional references about configuring QoS.

This chapter includes the following sections:

Preparing to Configure QoS

Before you begin to configure QoS for the Layer 3 queues, you should have a good understanding of the Cisco 7200 series architecture and how the router processes ATM traffic. In particular, you need to understand when Layer 3 queues are activated. This information is provided in Chapter 2, "Cisco 7200 Series Architecture and Design for ATM Traffic Management."

You also need to evaluate your network traffic and assess your business needs to establish the criteria for appropriate service policies before you configure QoS on your network. A description of the kinds of information that you should acquire and the tasks that you should perform are shown in Chapter 4, "Preparing to Configure ATM Traffic Management and QoS Features."

Architecture Overview

This section describes some of the important characteristics about the architecture used by the Cisco 7200 router during ATM processing that might be helpful for you to review as you prepare to configure QoS service policies:

- The Cisco 7200 series router activates Layer 3 queues whenever congestion builds on an egress interface and outbound traffic can not be processed to the transmit ring, or onto the hardware queue located on the ATM port adapter. Traffic shaping is frequently a cause for congestion on the egress ATM interface.
- For PA-A3 and PA-A6 ATM port adapters, there is a hold queue for each PVC that is configured on that interface. This environment provides more control and prevents any single over-subscribed PVC from starving other PVCs for transmission resources. This queue is called the per-VC queue.
- With the exception of process-switched packets, whenever entries are available for packets on the transmit ring, packets go directly to the transmit ring on the NPE and onto the FIFO hardware queue located on the ATM port adapter. Process-switched packets always enqueue to the Layer 3 queue first, before being placed onto the transmit ring, regardless of availability on the ring.
- QoS service policies only begin to apply to CEF-switched and fastswitched packets when there is congestion on the ATM port adapter and the transmit ring is full.

IP to ATM CoS Overview

The Cisco IOS software classifies QoS features into the following categories:

- Classification and Marking
- Congestion Avoidance
- Congestion Management
- Traffic Shaping and Policing
- Signaling
- Link Efficiency Mechanisms

IP to ATM Class of Service (CoS) refers to a subset of the overall QoS features available on the Cisco 7200 series router that enable you to specify queueing service policies on a per-VC basis. IP to ATM CoS identifies certain QoS features that can be specifically applied at a more discreet, per-VC level for PVCs on the PA-A3 and PA-A6 ATM port adapters:

- Congestion avoidance policies determine the drop behavior in a queue. For congestion avoidance, you can configure Weighted Random Early Detection (WRED) on a per-VC basis on the PA-A3 and PA-A6 ATM port adapters. The default behavior is tail drop.
- Congestion management policies determine the order of the queue. For congestion management, you can configure Low Latency Queueing (LLQ) or Class-Based Weighted Fair Queueing (CBWFQ) on a per-VC basis on the PA-A3 and PA-A6 ATM port adapters. The default congestion management policy is FIFO.

It is important to remember that queueing occurs on the outbound path only. Therefore, you can configure a service policy that specifies a non-default queueing strategy, such as WRED, CBWFQ, or LLQ, for outbound traffic only on a PVC. If you attempt to configure WRED, CBWFQ, or LLQ as an inbound policy, the command will be rejected and a message similar to the following appears:

```
*Jun  4 07:27:17.210:  CBWFQ : Can be enabled as an output feature only
```

Although only WRED, CBWFQ, and LLQ are supported on a per-VC, outbound basis, you can still use other QoS features to classify and mark different IP traffic on the inbound path in combination with implementing IP to ATM CoS features at the PVC. This chapter focuses only on guidelines for the IP to ATM CoS features.

Understanding the Queue Limit

The queue limit is an important concept to understand when configuring IP to ATM CoS features, and there are several aspects to consider. The hold queue limit specifies the maximum number of packets that can be held in the Layer 3 queue. The queue limit is called the *queue depth* also. For some port adapters, this is an interface level queue that all VCs share. For the PA-A3 and PA-A6 ATM port adapters, it is a per-VC queue, and the interface queue is not used.

The type of queueing strategy implemented on the VC determines the default queue limit size, and also determines which command you might use to customize the queue depth. When the number of packets in the queue reaches the queue limit for that VC, then the router initiates a drop policy. By default, there technically is not a congestion avoidance policy. The default behavior is tail drop, which occurs when the queue is full. When the VC uses a tail drop policy, the last packet in is the first packet dropped, as long as there is not any available space in the queue.

To minimize tail drop, you can configure WRED as an alternative congestion avoidance policy to implement drop probabilities among different classified flows of traffic. WRED provides intelligent dropping. For more information, see the "Configuring WRED" section on page 6-6.

There are two different types of queue limits that you can tune for the PA-A3 and PA-A6 ATM port adapters, which are used for different types of queueing strategies:

- Per-VC queue—This Layer 3 queue applies at the VC level when FIFO queueing is being used by the VC. The default size is 40 packets, but you can tune the size of the queue using the **vc-hold-queue** ATM VC configuration command.
- Class queue—This Layer 3 queue applies at the class level when CBWFQ is being used by the VC. The default size is 64 packets, but you can tune the size of the queue using the **queue-limit** policy-map class configuration command.

  CBWFQ creates a queue for every class for which a class map is defined. Each class has a queue limit associated with it, which specifies the maximum number of packets that can enqueue there. Packets that satisfy the match criteria for a class accumulate in the queue reserved for the class until they are sent, which occurs when the fair queueing process services the queue. When the class queue reaches the maximum packet threshold, enqueueing of any further packets to the class queue causes tail drop. If Weighted Random Early Detection (WRED) is configured for the class policy, intelligent packet drop takes effect before the queue limit is reached, while the average queue depth is between the minimum and maximum thresholds.

Configuring the Queue Limits

You can configure a per-VC queue limit when you use the FIFO queueing strategy at a VC, or you can configure a class queue limit at the VC if you are configuring a class-based policy. The default queue limit varies by the type of queueing policy that you configure.

For LLQ, the original default limit was 64. The queue limit is not configurable for LLQ. However, as of Cisco IOS release 12.1(3)T, the queue limit automatically adjusts to the configured bandwidth to accommodate packet bursts. For more information, see the *Configuring Burst Size in Low Latency Queueing* feature documentation.

Configuring the FIFO Per-VC Hold Queue Limit

For FIFO, the default maximum number of packets that can be in the per-VC queue is 40 packets. When using the default FIFO policy on a PVC, you can modify the queue limit associated with each VC using the **vc-hold-queue** ATM VC configuration command. The possible queue limits are 5 to 1024 packets.

FIFO Per-VC Hold Queue Limit Configuration Example

The following example specifies that PVC 0/100 can hold up to 50 packets before activating its drop policy during FIFO queueing:

```
Router(config)# interface atm3/0.1
Router(config-if)# pvc 0/100
Router(config-if-atm-vc)# vc-hold-queue 50
```

Verifying the Per-VC Hold Queue Limit

To verify the size of the per-VC hold queue on an ATM interface, use the **show queueing interface atm** command and observe the value of the “Output queue” field. The following example shows that the total possible per-VC queue depth is 50 packets, and no packets are currently in the queue (shown by “Output queue 0/50”):

```
Router# show queueing interface atm 3/0
  Interface ATM3/0 VC 0/100
  Queueing strategy: fifo
  Output queue 0/50, 0 drops per VC
[text omitted]
```

Configuring the Class Queue Limit

For CBWFQ, the default limit is 64 packets. When using a CBWFQ policy on a PVC, you can modify the queue limit associated with a class using the **queue-limit** policy-map class configuration command. The possible queue limits are 1 to 64 packets.

When you configure CBWFQ, the per-VC hold queue limit does not apply to the VC.

Class Queue Limit Configuration Example

The following example configures a policy map called “fairq” for the default class. The queue limit for this class is 50 packets:

```
Router(config)# policy-map fairq
Router(config-pmap)# class class-default
Router(config-pmap-c)# fair-queue
Router(config-pmap-c)# queue-limit 50
```

Verifying the Class Queue Limit

Once you have configured CBWFQ and attached the policy to a VC, you can use the **show queueing interface atm** command to verify the queueing strategy and observe the value of the “threshold” field to verify the limit for the class queue. The following example shows that the WFQ class limit on VC 10/32 is 50 packets:

```
Router# show queueing interface atm 2/0.100032
     Interface ATM2/0.100032 VC 10/32
     Queueing strategy: weighted fair
     Total output drops per VC: 1539
     Output queue: 0/512/50/1539 (size/max total/threshold/drops)
        Conversations  0/37/128 (active/max active/max total)
        Reserved Conversations 0/0 (allocated/max allocated)
```

Using MQC to Configure and Apply QoS Service Policies

The Modular QoS CLI (MQC) is a command-line interface (CLI) structure that allows you to create service polices and attach these policies to interfaces, subinterfaces, and ATM or Frame Relay virtual circuits (VCs). A traffic policy contains a traffic class and one or more QoS features. A traffic class is used to classify traffic, while the QoS features in the service policy determine how to treat the classified traffic.

For ATM port adapters, you can use MQC to create traffic policies and attach these policies at the PVC level, or at the interface level. For the PA-A3 and PA-A6 ATM port adapters, you should attach the policies to the VC, not the interface.

To create QoS classes and configure policies using MQC on the PA-A3 and PA-A6 ATM port adapters, complete the following basic steps:

Step 1 Define a traffic class using the **class-map** command.

Step 2 Create a traffic policy and associate the traffic class with one or more QoS features using the **policy-map** command.

Step 3 Attach the traffic policy to the PVC using the **service-policy** command.

For examples using MQC configuration, see Chapter 8, "ATM Traffic Management Case Studies and Configuration Examples." You can also find MQC configuration examples in the *Cisco IOS Quality of Service Solutions Configuration Guide* and *Cisco IOS Quality of Service Solutions Command Reference*.

Configuring WRED

Weighted Random Early Detection (WRED) is the Cisco Systems implementation of the Random Early Detection (RED) algorithm, which combines the capabilities of the RED algorithm with consideration of IP precedence. WRED flows are classified by different IP precedence levels.

WRED generally drops packets selectively based on IP precedence. Packets with a higher IP precedence are less likely to be dropped than packets with a lower precedence. Thus, the higher the priority of a packet, the higher the probability that the packet will be delivered.

WRED makes early detection of congestion possible and provides for multiple classes of traffic. It also protects against global synchronization and is therefore most useful in networks that transmit a large amount of TCP traffic. WRED reduces the chances of tail drop by selectively dropping packets when the output interface begins to show signs of congestion. By dropping some packets early rather than waiting until the queue is full, WRED avoids dropping large numbers of packets at once and minimizes the occurrence of global synchronization. Thus, WRED allows the bandwidth to be more efficiently used at all times.

Although WRED is an alternative method of congestion avoidance, it does not necessarily eliminate tail drop from occurring. WRED is a more proactive implementation than tail drop, because the WRED algorithm applies before the queue becomes full. If the queue limit is reached on the VC, then the tail drop strategy always applies. Therefore, you need to be careful about how you configure your queue limits and the maximum threshold for WRED to be sure that WRED can be properly activated before the queue limit is reached. If you configure WRED within a class policy, then the queue limit for the class applies.

For more information about global synchronization, see the "Congestion on an ATM Network" section on page 1-10. For more information about WRED, refer to the "Congestion Avoidance" section of the *Cisco IOS Quality of Service Solutions Configuration Guide*. You also can refer to the *IP to ATM Class of Service Phase 1 Design Guide* for more in-depth analysis of WRED implementation considerations. However, be aware that much of the discussion in the *IP to ATM Class of Service Phase I Design Guide* is based on the Cisco 7500 series router.

WRED Configuration Guidelines

When you configure WRED on a VC for the PA-A3 and PA-A6 ATM port adapters, consider the following guidelines:

- Be sure to enable Cisco Express Forwarding (CEF) switching, which is a requirement for WRED.
- Use WRED on VCs where you expect high congestion to occur, or that will be transmitting a high volume of Transmission Control Protocol (TCP) traffic.
- You can configure both WRED and a fancy queueing mechanism like CBWFQ or LLQ.
- Be aware that by default, non-IP traffic is treated with a precedence of zero, which means that your non-IP traffic will be dropped more often than your TCP/IP traffic.
- To transport IP best-effort traffic on the IP backbone, you need to implement a consistent policy for use of precedence values for different types of traffic throughout the network. In particular, you should perform precedence marking of IP traffic on the edge of the network (for example, marking of incoming IP precedence through the Cisco IOS Committed Access Rate [CAR] feature or through policy routing).
- WRED is supported at the interface level. However, for the PA-A3 and PA-A6 ATM port adapters, you should configure WRED at the per-VC level using the **random-detect** command. You can use a QoS policy map to apply WRED to the VC.
- The queue limit defines the maximum number of packets that the Layer 3 queues can store at any time. When the mean queue depth is between the minimum and maximum thresholds, WRED applies.
- If you change the queue limit but you are using the default WRED settings, then the maximum threshold automatically adjusts to the configured queue limit. However, if you manually configure the max-threshold, then you will lose the benefit of dynamic adjustment.

  The queue limit should be equal to or larger than the WRED max-threshold. If the per-VC queue limit is smaller, then the WRED mechanism can not be fully implemented on the VC because tail drop is enforced when the number of packets in the queue reaches the queue limit.
- Use the default parameters for WRED. The default WRED settings are very robust and automatically implement the following considerations:
  - The experience developed in the Internet research community on RED parameter setting
  - Configuration of related parameters (such as shaping parameters of the ATM VC on which WRED is run)
  - A different discard profile per precedence (the higher the precedence, the better the default corresponding service)
- The default values allocate the same max-thresholds and the same mark-probability to all the precedences. However, the default min-threshold is different for every precedence. The higher the precedence, the higher the min-thresholds. Consequently the default WRED configuration offers an increasingly better service to higher precedences.

⚠ **Caution** Because of the dynamic nature of RED and WRED and their complex interactions with transport-level flow control mechanisms (such as TCP flow control), fine-tuning WRED to achieve specific IP service differentiation objectives in particular operating conditions is a delicate exercise and great caution is recommended. We recommend that you start operations or testing with the default WRED settings (or from configurations close to the WRED default settings) and fine-tune from there. Modifications to WRED parameters should be tested and validated under a vast range of network conditions before being deployed in a large network.

WRED Configuration Example on an ATM PVC

To configure WRED using MQC, complete the following steps:

Step 1 From global configuration mode, enable IP CEF:

```
Router(config)# ip cef
```

Step 2 From global configuration mode, create the policy and configure WRED. The following example creates a policy named "atm_wred." The **bandwidth** command implements WFQ, and the **random-detect** command without any other parameters enables WRED using the default weights and precedence for the class named "mytest" in the policy:

```
Router(config)# policy-map atm_wred
Router(config-pmap)# class mytest
Router(config-pmap-c)# bandwidth 64
Router(config-pmap-c)# random-detect
Router(config-pmap-c)# exit
```

Step 3 From global configuration mode, create the PVC at the interface and apply the service policy. The following example configures PVC 1/120 at an ATM point-to-point interface and applies the service policy named "atm_wred" for traffic outbound on the PVC:

```
Router(config)# interface ATM1/0.20 point-to-point
Router(config-if)# ip address 10.20.20.21 255.255.255.0
Router(config-if)# pvc 1/120
Router(config-if-atm-vc)# vbr-nrt 150 100 120
Router(config-if-atm-vc)# service-policy output atm_wred
```

For more information about configuring WRED, refer to the *Cisco IOS Quality of Service Solutions Configuration Guide*.

Monitoring WRED Status on a VC

To see how many packets have been selectively dropped by the WRED algorithm on a per-VC basis, you can use the **show policy-map interface atm** command and observe the value of the "Random drop" counters. The following example shows that no drops are present on VC 1/120:

```
Router# show policy-map interface atm 1/0.20 out
 ATM1/0.20: VC 1/120 -

  Service-policy output: atm_wred

    Class-map: mytest (match-all)
      169 packets, 191676 bytes
      5 minute offered rate 0 bps, drop rate 0 bps
      Match: access-group 121
      Queueing
        Output Queue: Conversation 25
        Bandwidth 64 (kbps)
        (pkts matched/bytes matched) 112/187344
```

```
           (depth/total drops/no-buffer drops) 0/0/0
            exponential weight: 9
            mean queue depth: 0

      class     Transmitted       Random drop      Tail drop    Minimum Maximum  Mark
                pkts/bytes        pkts/bytes       pkts/bytes    thresh  thresh prob
          0      112/187344           0/0              0/0           20      40 1/10
          1        0/0                0/0              0/0           22      40 1/10
          2        0/0                0/0              0/0           24      40 1/10
          3        0/0                0/0              0/0           26      40 1/10
          4        0/0                0/0              0/0           28      40 1/10
          5        0/0                0/0              0/0           30      40 1/10
          6       63/4788             0/0              0/0           32      40 1/10
          7        0/0                0/0              0/0           34      40 1/10
       rsvp        0/0                0/0              0/0           36      40 1/10
```

Configuring CBWFQ

Native (flow-based) WFQ assigns a weight to each conversation, and then schedules the transmit time for each packet of the different flows. The weight is a function of the IP precedence of each flow, and the scheduling time depends on the packet size.

CBWFQ assigns a weight to each configured class instead of each flow. The bandwidth you assign to a class is used to calculate the weight of that class. More precisely, the weight is a function of the interface bandwidth divided by the class bandwidth. Therefore, the bigger the bandwidth parameter, the smaller the weight.

Without LLQ, CBWFQ provides weighted fair queueing based on defined classes with no strict priority queue available for real-time traffic. CBWFQ allows you to define traffic classes and then assign characteristics to those classes. For example, you can designate the minimum bandwidth delivered to a class during congestion.

For CBWFQ, the weight for a packet belonging to a specific class is derived from the bandwidth you assign to the class when you configure it. Therefore, the bandwidth assigned to the packets of a class determines the order in which packets are sent. All packets are serviced fairly based on weight; no class of packets may be granted strict priority.

This scheme poses problems for voice traffic that is largely intolerant of delay, especially variation in delay. For voice traffic, variations in delay introduce irregularities of transmission manifesting as jitter in the heard conversation. To overcome this limitation for voice traffic, you can use LLQ with CBWFQ.

CBWFQ Configuration Guidelines

When you configure CBWFQ on a VC for the PA-A3 and PA-A6 ATM port adapters, consider the following guidelines:

- The size of the transmit ring limit determines how quickly the Layer 3 queue is activated. Therefore, when you plan to implement WFQ, you should reduce the transmit ring limit. For more information, see Chapter 7, “Configuring the Ring Limits on the PA-A3 and PA-A6 ATM Port Adapters.”
- Implement CBWFQ on slow PVCs if you do not want bulk traffic to impact the transmission of smaller packet traffic.
- Modify the queue limit for each class using the **queue-limit** policy-map class configuration command. The possible queue limits are 1 to 64 packets.
- You can configure WRED with CBWFQ as an alternative drop strategy.

- Because CBWFQ provides a minimum bandwidth guarantee, you can only apply CBWFQ to VCs with classes of service other than UBR and UBR+. These service categories are best-effort classes that do not guarantee a minimum bandwidth.
- The PA-A3 and PA-A6 ATM port adapters do not support native, flow-based WFQ configured directly on an interface using the **fair-queue** command. You need to configure WFQ within the default class using a policy map to implement CBWFQ for the PA-A3 and PA-A6 ATM port adapters.
- Use class maps to classify and assign weights to traffic. Classification parameters and class maps are defined at the same place.
- After defining the classification parameters, configure a policy map to apply traffic parameters to these classified flows.
- After you configure the traffic parameters for each class, apply CBWFQ on a VC-basis using the **service-policy output** ATM VC configuration command.
- Traffic that does not match one of the defined class maps is assigned a default class map (class default) that you define in the policy map. The parameters configured under this default class apply to all non-classified traffic.

CBWFQ Configuration Example

To configure CBWFQ using MQC, complete the following steps:

Step 1 From global configuration mode, create the policy and configure fair queueing. The following example creates a policy named “cbwfq” for the class called “mytest.” The **bandwidth** command implements CBWFQ for the class:

```
Router(config)# policy-map cbwfq
Router(config-pmap)# class mytest
Router(config-pmap-c)# bandwidth 256
Router(config-pmap-c)# end
```

Step 2 Beginning in global configuration mode, create the interface, create the PVC, and apply the service policy. The following example configures PVC 0/101 at an ATM point-to-point interface and applies the service policy named “cbwfq” for traffic outbound on the PVC:

```
Router(config)# interface atm 4/0.11 point-to-point
Router(config-subif)# ip address 10.10.10.1 255.255.255.0
Router(config-subif)# pvc 0/101
Router(config-if-atm-vc)# vbr-nrt 2048 1024 96
Router(config-if-atm-vc)# service-policy output cbwfq
```

For more information about configuring CBWFQ, refer to the *Cisco IOS Quality of Service Solutions Configuration Guide*.

Monitoring CBWFQ Status on a VC

To monitor the status of CBWFQ on a VC, complete the following steps:

Step 1 From privileged EXEC mode, run the **show queue** command for the VC that you want to see:

```
Router# show queue atm 4/0.11 vc 0/101
  Interface ATM4/0.11 VC 0/101
  Queueing strategy: weighted fair
  Output queue: 56/512/64/38858 (size/max total/threshold/drops)
     Conversations  2/2/64 (active/max active/max total)
     Reserved Conversations 1/1 (allocated/max allocated)
     Available Bandwidth 512 kilobits/sec

  (depth/weight/total drops/no-buffer drops/interleaves) 1/228/9284/0/0
  Conversation 73, linktype: ip, length: 994
  source: 10.0.0.2, destination: 10.10.10.2, id: 0x0000, ttl: 63, prot: 255

  (depth/weight/total drops/no-buffer drops/interleaves) 55/32384/29574/0/0
  Conversation 44, linktype: ip, length: 994
  source: 10.0.0.2, destination: 10.10.10.3, id: 0x0000, ttl: 63, prot: 255
```

Step 2 To verify how many packets are currently in the hold queue across all conversations, observe the value of the “Output queue” size field. The following example shows that the per-VC queue is 56 packets for VC 0/101:

```
Router# show queue atm 4/0.11 vc 0/101
  Interface ATM4/0.11 VC 0/101
  Queueing strategy: weighted fair
  Output queue: 56/512/64/38858 (size/max total/threshold/drops)
     Conversations  2/2/64 (active/max active/max total)
[text omitted]
```

Step 3 To verify how many packets are in the queue for each conversation flow, observe the value of the “depth” fields, as shown in the following output excerpt:

```
[text omitted]
(depth/weight/total drops/no-buffer drops/interleaves) 1/228/9284/0/0
  Conversation 73, linktype: ip, length: 994
  source: 10.0.0.2, destination: 10.10.10.2, id: 0x0000, ttl: 63, prot: 255

  (depth/weight/total drops/no-buffer drops/interleaves) 55/32384/29574/0/0
  Conversation 44, linktype: ip, length: 994
  source: 10.0.0.2, destination: 10.10.10.3, id: 0x0000, ttl: 63, prot: 255
[text omitted]
```

Step 4 To verify how many packets you can queue for each conversation, observe the value of the “Output queue” threshold field as shown in the following output excerpt. Notice that for the WFQ queueing strategy, the number of possible packets in the hold queue is 64 (the default). The total number of packets that all conversations can queue is 512:

```
Router# show queue atm 4/0.11 vc 0/101
  Interface ATM4/0.11 VC 0/101
  Queueing strategy: weighted fair
  Output queue: 56/512/64/38858 (size/max total/threshold/drops)
[text omitted]
```

Step 5 To verify information about the classes for the output policy on a PVC, run the **show policy-map interface** command:

```
Router# show policy-map interface atm 4/0.11 vc 0/101
 ATM4/0.11: VC 0/101 -

  Service-policy output: cbwfq

    Class-map: mytest (match-all)
      153656 packets, 152734064 bytes
      30 second offered rate 230000 bps, drop rate 0 bps
      Match:  precedence 6
      Queueing
        Output Queue: Conversation 73
        Bandwidth 256 (kbps) Max Threshold 64 (packets)
        (pkts matched/bytes matched) 146307/145429158
        (depth/total drops/no-buffer drops) 0/9284/0

    Class-map: class-default (match-any)
      257250 packets, 255704576 bytes
      30 second offered rate 224000 bps, drop rate 93000 bps
      Match: any
```

Configuring LLQ

The Low Latency Queueing feature brings strict priority queueing to CBWFQ. Configured by the **priority** command, strict priority queueing gives delay-sensitive data, such as voice, preferential treatment over other traffic. With this feature, delay-sensitive data is sent first, before packets in other queues are serviced.

LLQ Configuration Guidelines

When you configure LLQ on a VC for the PA-A3 and PA-A6 ATM port adapters, consider the following guidelines:

- The size of the transmit ring limit determines how quickly the Layer 3 queue is activated. Therefore, when you plan to implement LLQ you should reduce the transmit ring limit. For more information, see Chapter 7, “Configuring the Ring Limits on the PA-A3 and PA-A6 ATM Port Adapters.”
- The **priority** command allows you to set up classes based on a variety of criteria (not just User Datagram Protocol [UDP] ports) and assign priority to them, and is available for use on serial interfaces and ATM permanent virtual circuits (PVCs). A similar command, **ip rtp priority**, allows you to define priority flows based only on UDP port numbers, but it is not available for ATM PVCs.
- Layer 2 encapsulations are accounted for in the amount of bandwidth specified with the **priority** command. However, the amount of bandwidth does not include other headers such as ATM cell tax overheads. You must also allow bandwidth for possible jitter introduced by the routers in the voice path.
- Use the **priority** command for Voice over IP (VoIP) on serial links and ATM PVCs.

Note In Cisco IOS Release 12.0(7)T and Cisco IOS Release 12.1, the **priority** command does not support VoIP over Frame Relay links. As of Cisco IOS Release 12.2 and later, the **priority** command is supported on Frame Relay links.

- You cannot configure the **priority** command with the **random-detect** command (for WRED), or with the the **queue-limit** command (to configure class queue depth). The **bandwidth** command and **priority** command are mutually exclusive.
- You can configure the **priority** command in multiple classes, but you should only use it for voice-like, constant bit rate (CBR) traffic.
- The functionality of LLQ has been extended to allow a configurable Committed Burst (Bc) size using the *Configuring Burst Size in Low Latency Queueing* feature. With this new functionality, the network can now accommodate temporary bursts of traffic and handle network traffic more efficiently.

LLQ Configuration Examples

The following example configures strict priority queueing with a guaranteed bandwidth of 50 kbps for the policy map named “llq”:

```
Router(config)# policy-map llq
Router(config-pmap)# class voice
Router(config-pmap-c)# priority 50
```

In this example, traffic classes named “voice” and “video” go into the high priority queue and get strict priority queueing over data traffic. However, voice traffic will be rate-limited to 50 Kbps and video traffic will be rate-limited to 100 Kbps. The classes will be individually rate-limited even if they go into the same queue:

```
Router(config)# policy-map llq
Router(config-pmap)# class voice
Router(config-pmap-c)# priority 50
Router(config-pmap-c)# exit
Router(config-pmap)# class video
Router(config-pmap-c)# priority 100
Router(config-pmap-c)# exit
Router(config-pmap)# class data
Router(config-pmap-c)# bandwidth 500
```

Monitoring LLQ Status on a VC

To monitor LLQ status on a VC, use the **show policy-map interface atm** command.

The following sample output was obtained from an ATM PVC with an SCR of 1024 Kbps. For LLQ, the queueing system adjusts the burst size as the value of the **priority** command changes:

```
Router# show policy-map interface atm 4/0.11 vc 0/101
 ATM4/0.11: VC 0/101 -

  Service-policy output: llq

    Class-map: data (match-all)
      79793 packets, 79314242 bytes
      30 second offered rate 254000 bps, drop rate 0 bps
      Match: ip precedence 0
    Queueing
      Output Queue: Conversation 73
      Bandwidth 500 (kbps) Max Threshold 64 (packets)
      (pkts matched/bytes matched) 38326/38096044
      (depth/total drops/no-buffer drops) 0/1693/0

  Class-map: voice (match-all)
    996 packets, 93624 bytes
    30 second offered rate 5000 bps, drop rate 0 bps
    Match: ip precedence 5
    Queueing
      Strict Priority
      Output Queue: Conversation 72
      Bandwidth 50 (kbps) Burst 1250 (Bytes)
      (pkts matched/bytes matched) 0/0
      (total drops/bytes drops) 0/0

  Class-map: video (match-all)
    1029 packets, 96726 bytes
    30 second offered rate 5000 bps, drop rate 0 bps
    Match: ip precedence 6
    Queueing
      Strict Priority
      Output Queue: Conversation 72
      Bandwidth 100 (kbps) Burst 2500 (Bytes)
      (pkts matched/bytes matched) 503/47282
      (total drops/bytes drops) 0/0

  Class-map: class-default (match-any)
    1 packets, 32 bytes
    30 second offered rate 0 bps, drop rate 0 bps
    Match: any
```

Monitoring QoS on the PA-A3 and PA-A6 ATM Port Adapters

To monitor per-VC drop counters on the PA-A3 and PA-A6 ATM port adapters, you need to use the **show queueing interface atm** command. Do not use the **show atm vc** command.

Related Documentation

The following table provides information about additional resources that you can read to learn more about some of the topics discussed in this chapter:

For more information about:	Refer to the following publications:
Cisco IOS QoS software commands	*Cisco IOS Quality of Service Solutions Command Reference*
Cisco IOS QoS software features	*Cisco IOS Quality of Service Solutions Configuration Guide*
Applying Service Policies for ATM	Where Do I Apply a QoS Service Policy on an ATM Interface? (TAC Tech Note)
Burst sizes and LLQ	Configuring Burst Size in Low Latency Queueing (Cisco IOS feature module)

For more information about:	Refer to the following publications:
CBWFQ an transmit ring limit relationship	Understanding Class Based Weighted Fair Queueing on ATM (TAC Tech Note)
Per-VC CBWFQ	Per-VC Class-Based, Weighted Fair Queueing (Per-VC CBWFQ) on the Cisco 7200, 3600, and 2600 Routers (TAC Tech Note)
QoS FAQs	QoS Frequently Asked Questions
WRED implementation and fine-tuning	*IP to ATM Class of Service Phase 1 Design Guide*

Next Steps

This chapter provides guidelines and information about queue limits and how to configure the IP to ATM CoS features that are supported on the PA-A3 and PA-A6 ATM port adapters, including WRED, CBWFQ, and LLQ.

To activate Layer 3 queues, you might need to optimize the size of the transmit ring for the ATM port adapter on the NPE/NSE. For more information, see Chapter 7, "Configuring the Ring Limits on the PA-A3 and PA-A6 ATM Port Adapters."

For in-depth examples and case studies of QoS configuration in an ATM network, see Chapter 8, "ATM Traffic Management Case Studies and Configuration Examples."