



Implementing a Wide Area Network

This chapter covers design considerations and recommendations for integrating your Cisco AVVID solution with a WAN.

WAN QoS Overview

A lower total cost of ownership is one of the most compelling reasons for migrating to a converged data, voice, and video network. While a converged network can lower overall costs of the enterprise communications infrastructure, solid planning and design is still required for a successful Cisco AVVID deployment. Nowhere is this fact more evident than when running VoIP over a Wide Area Network (WAN).

As stated in Chapter 1, “Overview,” three basic tools must be used on every portion of the IP network to provide an environment that can ensure voice quality over the network:

- Classification
- Queuing
- Network provisioning

When the low bandwidths and slow link speeds of a WAN are introduced into a Cisco AVVID design, you must also use several additional QoS tools:

- Link Fragmentation and Interleaving (LFI)
- Traffic shaping
- Call admission control

All of these tools, plus several others, are described in the following sections.

Classification

Classification is the method by which certain traffic types are classified, or marked, as having unique handling requirements. These requirements might be a minimum required amount of bandwidth or a low tolerance for latency. This classification can be signaled to the network elements via a tag included in the IP Precedence or Differentiated Services Code Point (DSCP), in Layer 2 schemes such as 802.1p, in the source and destination IP addresses, or in the implicit characteristics of the data itself, such as the traffic type using the Real-time Transport Protocol (RTP) and a defined port range.

In the recommended Cisco AVVID QoS design model, classification is done at both Layer 2 and Layer 3 on the IP phone. In this model, the phone is the "edge" of the managed network, and it sets the Layer 2 802.1p CoS value to 5 and the Layer 3 IP Precedence value to 5 or the DSCP value to EF. For more details on classification, see Chapter 2, "Connecting IP Phones."

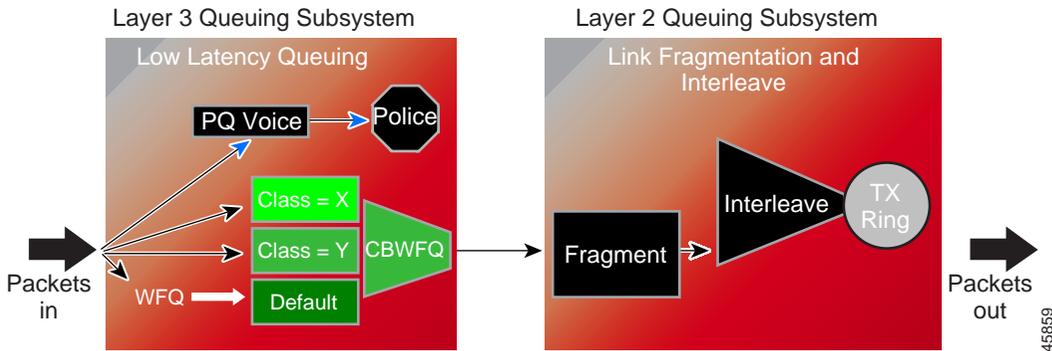
Queuing

As was discussed in previous chapters, interface queuing is one of the most important mechanisms for ensuring voice quality within a data network. This is even more vital in the WAN because many traffic flows are contending for a very limited amount of network resources. Once traffic has been classified, the flow can be placed into an interface egress queue that meets its handling requirements. Voice over IP, because of its extremely low tolerance for packet loss and delay, should be placed into a Priority Queue (PQ). However, other traffic types may have specific bandwidth and delay characteristics as well. These requirements are addressed with the Low-Latency Queuing (LLQ) feature in Cisco IOS.

LLQ combines the use of a PQ with a class-based weighted fair queuing scheme. Classes are defined with classification admission schemes. Traffic flows have access to either the PQ, one of the class-based queues, or a default weighted fair queue. LLQ, the recommended queuing scheme for all low-speed links, allows up to 64 traffic classes with the ability to specify such parameters as priority queuing behavior for voice, a minimum bandwidth for Systems Network Architecture (SNA) data, and Cisco AVVID control protocols and weighted fair queuing for other traffic types.

As depicted in Figure 5-1, when a Priority Queuing class is configured, the PQ has direct access to the transmit (TX) ring. This is, of course, unless interleaving is configured, in which case interleaving occurs prior to placing the PQ traffic onto the TX-ring.

Figure 5-1 Packet Flow with Priority Queuing



The maximum configured bandwidth in the PQs and class-based queues cannot exceed the minimum available amount of bandwidth on the WAN connection. A practical example is a Frame Relay LLQ with a Committed Information Rate (CIR) of 128 kbps. If the PQ for VoIP is configured for 64 kbps and both the SNA and Cisco AVVID control protocol class-based queues are configured for 20 kbps and 10 kbps, respectively, the total configured queue bandwidth is 94 kbps. Cisco IOS defaults to a minimum CIR (**mincir**) value of CIR/2. The **mincir** value is the transmit value a Frame Relay router will "rate down" to when Backward Explicit Congestion Notifications (BECNs) are received. In this example, the **mincir** value is 64 kbps and is lower than the configured bandwidth of the combined queues. For LLQ to work in this example, a **mincir** value of 128 kbps should be configured.

Link Fragmentation and Interleaving

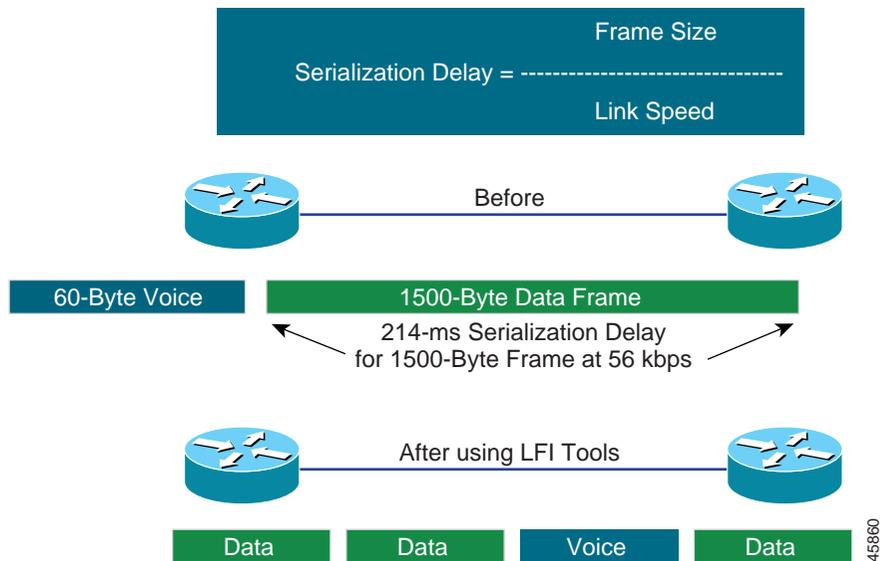
For low-speed WAN connections (in practice, those with a clocking speed of 768 kbps or below), it is necessary to provide a mechanism for Link Fragmentation and Interleaving (LFI). A data frame can be sent to the physical wire only at the serialization rate of the interface. This serialization rate is the size of the frame divided by the clocking speed of the interface. For example, a 1500-byte frame takes 214 ms to serialize on a 56-kbps circuit. If a delay-sensitive voice packet is behind a large data packet in the egress interface queue, the end-to-end delay budget of 150-200 ms could be exceeded. In addition, even relatively small frames can adversely affect overall voice quality by simply increasing the jitter to a value greater than the size of the adaptive jitter buffer at the receiver. Table 5-1 shows the serialization delay for various frame sizes and link speeds.

Table 5-1 Serialization Delay

Link Speed	Frame Size (Bytes)					
	64	128	256	512	1024	1500
56 kbps	9 ms	18 ms	36 ms	72 ms	144 ms	214 ms
64 kbps	8 ms	16 ms	32 ms	64 ms	128 ms	187 ms
128 kbps	4 ms	8 ms	16 ms	32 ms	64 ms	93 ms
256 kbps	2 ms	4 ms	8 ms	16 ms	32 ms	46 ms
512 kbps	1 ms	2 ms	4 ms	8 ms	16 ms	23 ms
768 kbps	0.640 ms	1.28 ms	2.56 ms	5.12 ms	10.4 ms	15 ms

LFI tools are used to fragment large data frames into regularly sized pieces and to interleave voice frames into the flow so that the end-to-end delay can be predicted accurately. This places bounds on jitter by preventing voice traffic from being delayed behind large data frames, as illustrated in Figure 5-2. The two techniques used for this are FRF.12 for Frame Relay and Multilink Point-to-Point Protocol (MLP) for point-to-point serial links.

Figure 5-2 Using LFI Tools to Reduce Frame Delay



A 10-ms blocking delay is the recommended target to use for setting fragmentation size. To calculate the recommended fragment size, divide the recommended 10 ms of delay by one byte of traffic at the provisioned line clocking speed, as follows:

$$\text{Fragment_Size} = (\text{Max_Allowed_Jitter} * \text{Link_Speed_in_kbps}) / 8$$

For example:

$$\text{Fragment_Size} = (10 \text{ ms} * 56) / 8 = 70 \text{ bytes}$$

Table 5-2 shows the recommended fragment size for various link speeds.

Table 5-2 Recommended Fragment Sizes

Link Speed	Recommended Fragment Size
56 kbps	70 bytes
64 kbps	80 bytes
128 kbps	160 bytes
256 kbps	320 bytes

Table 5-2 Recommended Fragment Sizes (continued)

Link Speed	Recommended Fragment Size
512 kbps	640 bytes
768 kbps	960 bytes

**Note**

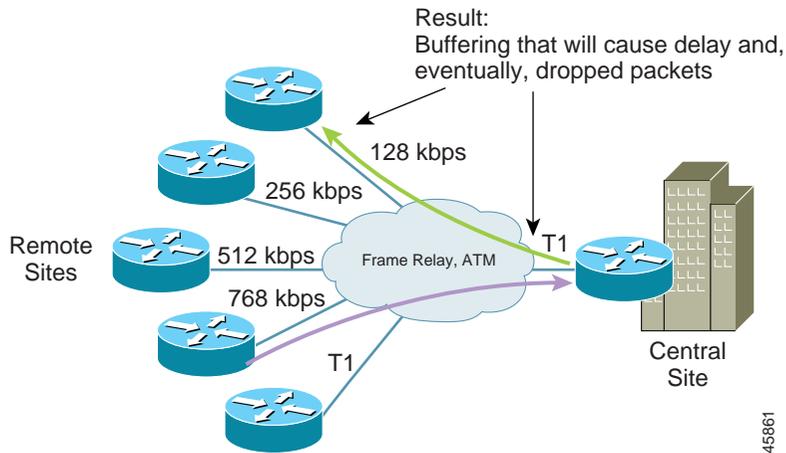
In Cisco IOS Release 12.1(5)T and later, MLP over ATM and Frame Relay are available to support LFI on ATM and ATM or Frame-Relay Interworking WANs.

Traffic Shaping

In ATM and Frame-Relay networks, where the physical access speed varies between two endpoints, traffic shaping is used to prevent excessive delay from congested network interface buffers caused by these speed mismatches. Traffic shaping is a tool that meters the transmit rate of frames from a source router to a destination router. This metering is typically done at a value that is lower than the line or circuit rate of the transmitting interface. The metering is done at this rate to account for the circuit speed mismatches that are common in current multiple-access, nonbroadcast networks.

Traffic leaving a high-speed interface such as a T1 line at a central site often terminates at a remote site that may have a much slower link speed (for example, 56 kbps). This is quite common and, in fact, has been one of the big selling points for Frame Relay. In Figure 5-3, the T1 interface on the router at the central site sends data out at a T1 rate even if the remote site has a clock rate of 56 kbps. This causes the frames to be buffered within the carrier Frame-Relay network, increasing variable delay, as illustrated in Figure 5-3. This same scenario can be applied in reverse. For example, the many remote sites, each with small WAN connections, when added together can oversubscribe the provisioned bandwidth or circuit speed at the central site.

Figure 5-3 Variable Delay Caused by Buffering

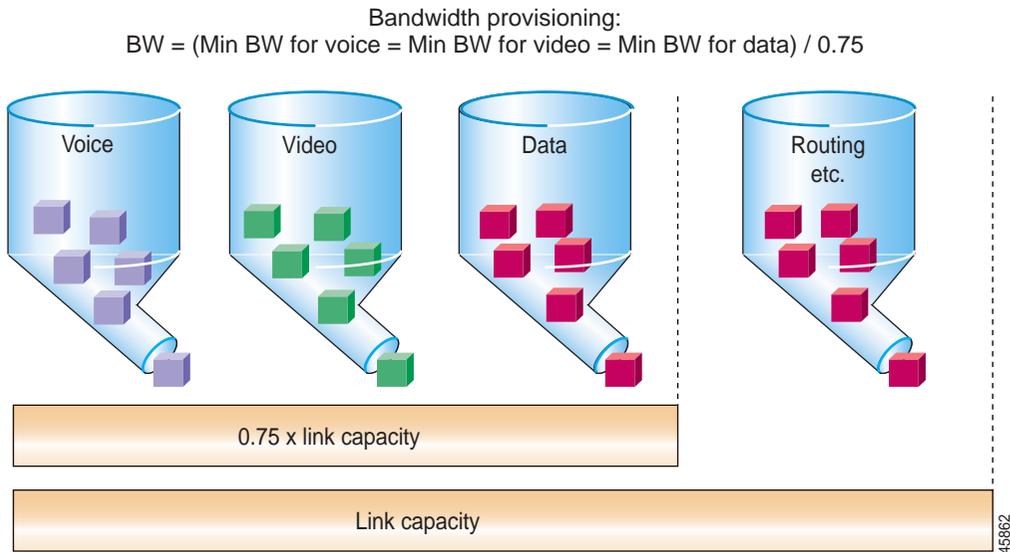


45861

Network Provisioning

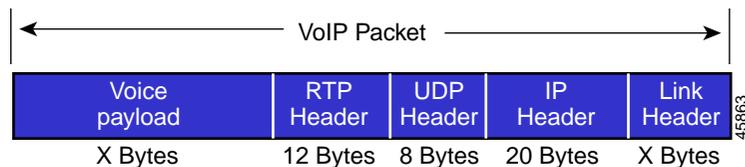
Properly provisioning the network bandwidth is a major component of designing a successful Cisco AVVID network. You can calculate the required bandwidth by adding the bandwidth requirements for each major application (for example, voice, video, and data). This sum then represents the minimum bandwidth requirement for any given link, and it should not exceed approximately 75% of the total available bandwidth for the link. This 75% rule assumes that some bandwidth is required for overhead traffic, such as routing and Layer 2 keepalives, as well as for additional applications such as e-mail and Hypertext Transfer Protocol (HTTP) traffic. Figure 5-4 illustrates this bandwidth provisioning process.

Figure 5-4 Provisioning Link Bandwidth



As illustrated in Figure 5-5, a VoIP packet consists of the payload, IP header, User Datagram Protocol (UDP) header, Real-time Transport Protocol (RTP) header, and Layer 2 Link header. At the default packetization rate of 20 ms, VoIP packets have a 160-byte payload for G.711 or a 20-byte payload for G.729. The IP header is 40 bytes, the UDP header is 8 bytes, and the RTP header is 12 bytes. The link header varies in size according to media.

Figure 5-5 Typical VoIP Packet



The bandwidth consumed by VoIP streams is calculated by adding the packet payload and all headers (in bits), then multiplying by the packet rate per second (default of 50 packets per second). Table 5-3 details the bandwidth per VoIP flow at a default packet rate of 50 packets per second (pps). This does not include Layer 2 header overhead and does not take into account any possible compression

schemes, such as compressed Real-time Transport Protocol (cRTP). You can use the Service Parameters menu in Cisco CallManager Administration to adjust the packet rate.

**Note**

While it is possible to configure the sampling rate above 30 ms, this usually results in very poor voice quality.

Table 5-3 Bandwidth Consumption for Voice Payload Only

CODEC	Sampling Rate	Voice Payload in Bytes	Packets per Second	Bandwidth per Conversation
G.711	20 ms	160	50	80 kbps
G.711	30 ms	240	33	53 kbps
G.729A	20 ms	20	50	24 kbps
G.729A	30 ms	30	33	16 kbps

A more accurate method for provisioning is to include the Layer 2 headers in the bandwidth calculations, as shown in Table 5-4.

Table 5-4 Bandwidth Consumption with Headers Included

CODEC	Ethernet 14 Bytes of Header	PPP 6 Bytes of Header	ATM 53-Byte Cells with a 48-Byte Payload	Frame-Relay 4 Bytes of Header
G.711 at 50 pps	85.6 kbps	82.4 kbps	106 kbps	81.6 kbps
G.711 at 33 pps	56.5 kbps	54.4 kbps	70 kbps	54 kbps
G.729A at 50 pps	29.6 kbps	26.4 kbps	42.4 kbps	25.6 kbps
G.729A at 33 pps	19.5 kbps	17.4 kbps	28 kbps	17 kbps

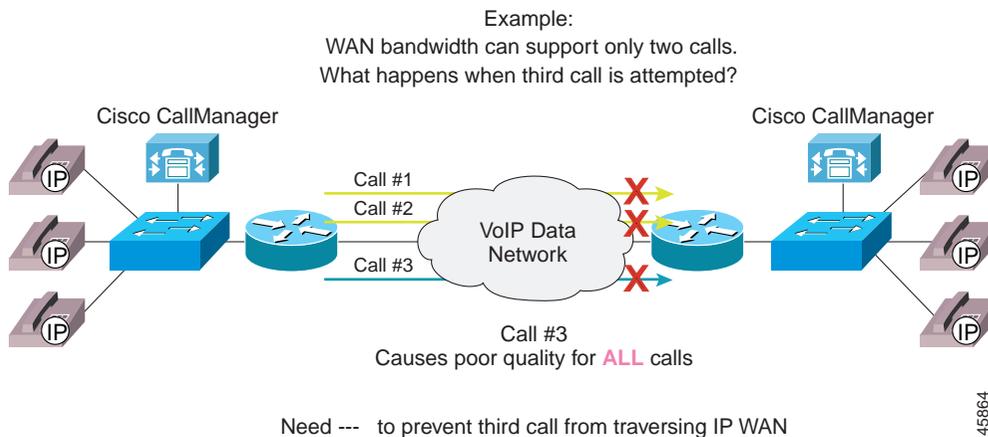
Call Admission Control

Call admission control is a mechanism for ensuring that voice flows do not exceed the maximum provisioned bandwidth allocated for voice conversations.

After doing the calculations to provision the network with the required bandwidth to support voice, data, and possibly video applications, it is important to ensure that voice does not oversubscribe the portion of the bandwidth allocated to it. While most QoS mechanisms are used to protect voice from data, call admission control is used to protect voice from voice. This is illustrated in Figure 5-6, which shows an environment where the network has been provisioned to support two concurrent voice calls. If a third voice call is allowed to proceed, the quality of all three calls is degraded. To prevent this degradation in voice quality, you can provision call admission control in Cisco CallManager to block the third call. For more information on call admission control, see the *Cisco IP Telephony Network Design Guide*, available at

http://www.cisco.com/univercd/cc/td/doc/product/voice/ip_tele/index.htm

Figure 5-6 Call Admission Control



Miscellaneous WAN QoS Tools

This section describes the following additional QoS tools, which can help ensure voice quality in WAN applications:

- VoIP Control Traffic
- TX-ring sizing
- Compressed voice codecs
- Compressed RTP (cRTP)
- Voice Activity Detection (VAD)

VoIP Control Traffic

When allocating bandwidth for the IP WAN, do not overlook the Cisco CallManager control traffic. In centralized call processing designs, the IP phones use a Transmission Control Protocol (TCP) control connection to communicate with Cisco CallManager. If there is not enough bandwidth provisioned for these small control connections, callers might be adversely affected.

An example where this comes into play is with the Delay-to-Dial-Tone (DTT) time. The IP phones communicate with Cisco CallManager via Skinny Station Protocol over TCP port 2001. When an IP phone goes off-hook, it "asks" Cisco CallManager what to do. Cisco CallManager instructs the IP phone to play dial tone. If this Skinny Protocol management and control traffic is dropped or delayed within the network, the user will not receive dial tone. This same logic applies to all signaling traffic for gateways and phones.

To ensure that this control and management traffic is marked as important (but not as important as voice), Access Control Lists (ACLs) are used to classify these streams on Layer 3 or 4 Catalyst 6000 switches at the central locations. Examples of these configurations are included in Chapter 3, "Designing a Campus." In the remote offices, a Cisco router might be the first Layer 3 or 4 device a packet

encounters before hitting the WAN. To ensure that these control connections are classified as important (but not as important as voice) access lists are used in the branch router, as illustrated in the following configuration example:

```
class-map VoIP-RTP
  match access-group 100
class-map VoIP-Control
  match access-group 101
!
policy-map QoS-Policy
  class VoIP-RTP
    priority 100
  class VoIP-Control
    bandwidth 8
  class class-default
    fair-queue
!
access-list 100 permit ip any any precedence 5
access-list 100 permit ip any any dscp ef
!
! Skinny Control Traffic - Not required with
! Cisco CallManager Release 3.0(5) and beyond.
access-list 101 permit tcp any host 10.1.10.20 range 2000 2002
!
! MGCP Control Traffic
access-list 101 permit udp any host 10.1.10.20 2427
access-list 101 permit tcp any host 10.1.10.20 2428
!
! H.323 Control Traffic
access-list 101 permit tcp any host 10.1.10.20 1720
access-list 101 permit tcp any host 10.1.10.20 range 11000 11999
```

TX-Ring Sizing

The TX-ring is the unprioritized FIFO buffer used to hold frames prior to transmission to drive link utilization to 100%. In the Cisco 7500 Route/Switch Processor (RSP), this is referred to as the TX-queue and can be modified using the **tx-queue-limit** command. The RSP is a very inefficient QoS platform, especially with regard to modifying the TX-queue parameters. The Cisco 7500 RSP TX-queue, which refers to the FIFO queue in MEM-D, has to copy the packet from MEM-D to the system buffers in DRAM and then back from the system buffers to MEM-D. The TX-ring is much more efficient than the TX-queue and is used instead of it on the Cisco 7500 VIP, 7200, 3600, 2600, and 1750 routers.

While fragmentation and interleaving reduces jitter, a large TX-ring value can increase jitter when link utilization approaches saturation. Because of this, TX-ring sizing is related to fragmentation size, as shown in Table 5-5.

**Note**

The sizing of the TX-ring buffer is measured in packets, not bits.

Table 5-5 TX-Ring Buffer Sizing

Link Speed (CIR) on Permanent Virtual Circuit	TX-Ring Buffer Sizing (Packets)
=< 128 kbps	5
192 kbps	6
256 kbps	7
512 kbps	14
768 kbps	21

On all Point-to-Point Protocol (PPP) and Multilink PPP (MLP) links, TX-ring buffer size is automatically configured, and you cannot change these default buffer values.

On Frame Relay links, the TX-ring is for the main interface, which all subinterfaces also use. The default TX-ring buffer size is 64 packets. You might need to change this setting when the subinterface is very small or there are many subinterfaces.

Table 5-6 summarizes TX-ring buffer sizing for various media.

Table 5-6 TX-Ring Buffer Sizing

Media	Default TX-Ring Buffer Sizing (Packets)
PPP	6
MLP	2
ATM	8192 (Must be changed for low-speed virtual circuits)
Frame Relay	64 (Per main T1 interface)

Compressed Voice Codecs

To utilize as much of the limited WAN bandwidth as possible, VoIP uses codecs (coding-decoding algorithms) to digitize analog voice samples. Many codecs, such as G.729, can compress a 64-kbps call down to 8 kbps. These types of codecs, termed low-bit-rate codecs, are commonly used for voice calls across the WAN.

Compressed RTP

Compressed RTP (cRTP) compresses the 40-byte header of a VoIP packet to approximately 2 to 4 bytes. Compressed RTP works on a link-by-link basis and is enabled on Cisco routers using the **ip rtp header-compression** command. Table 5-7 summarizes the bandwidth calculations for cRTP.



Note

cRTP is currently supported only for leased lines and Frame Relay. Cisco IOS Release 12.1(2)T, which greatly enhances performance over these platforms, is the minimum recommended system software for scalable cRTP.

Table 5-7 Compressed RTP Bandwidth Calculations

Codec	PPP 6 Bytes of Header	ATM 53-Byte Cells with a 48-Byte Payload	Frame Relay 4 Bytes of Header
G.711 at 50 pps	68 kbps	N/A	67 kbps
G.711 at 33 pps	44 kbps	N/A	44 kbps
G.729A at 50 pps	12 kbps	N/A	11.2 kbps
G.729A at 33 pps	8 kbps	N/A	7.4 kbps

Voice Activity Detection

Voice Activity Detection (VAD) takes advantage of the fact that, in most conversations, only one party is talking at a time. The VAD algorithm in the VoIP software examines the voice conversation, looking for these gaps in conversation. When a gap is discovered, no packets are sent, and the WAN bandwidth can be recovered for use by data applications. It is recommended you always turn VAD *off* systemwide.



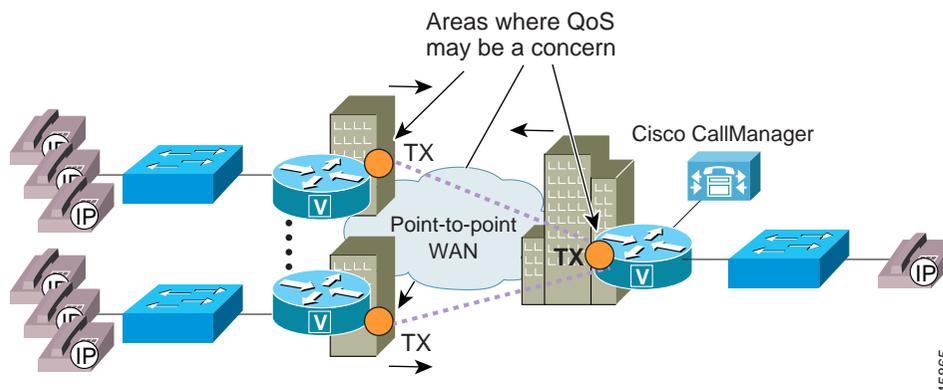
Note

In environments that have a large amount of inherent delay, VAD can sometimes cause more voice quality issues than are justified by the bandwidth recovered. You should examine these issues on a case-by-case basis. However, when troubleshooting clipping at the beginning of conversations in a Cisco AVVID network, it is advisable to disable Silence Suppression first.

Point-to-Point WAN

Point-to-point WANs, while not as popular as in the past, are still one of the most common types of networks in use today. Figure 5-7 shows the general model for point-to-point WANs described in this guide.

Figure 5-7 General Model for a Point-to-Point WAN



When designing a point-to-point WAN for a Cisco AVVID network, keep the following recommendations in mind:

- Cisco IOS Release 12.1(3)T is the minimum recommended release for a point-to-point WAN.
- Use Link Fragmentation and Interleaving (LFI) techniques on all WAN connections with speeds below 768 kbps.
- Use Low-Latency Queuing (LLQ) with a priority queue for VoIP bearer streams and a class queue for VoIP control sessions.
- Call admission control is required when the number of calls across the WAN can oversubscribe the allocated VoIP bandwidth.

The following sections explain the QoS issues for this type of configuration.

LFI on Point-to-Point WANs

If the clocking speed of the connection is below 768 kbps, LFI must be used. Multilink PPP (MLP) instead of PPP is required on all point-to-point links where LFI is needed. To enable LFI on point-to-point WANs, use the Cisco IOS command set for MLP.

**Note**

When using MLP, fragmentation size is configured using the maximum acceptable delay in queue, which is 10 ms. In addition, the TX-ring is statically configured at a value of 2 packets.

The following example illustrates the commands used for this type of configuration:

```
interface Multilink1
  ip address 10.1.61.1 255.255.255.0
  ip tcp header-compression iphc-format
  no ip mroute-cache
  load-interval 30
  service-policy output QoS-Policy
  ppp multilink
  ppp multilink fragment-delay 10
  ppp multilink interleave
  multilink-group 1
  ip rtp header-compression iphc-format
!
interface Serial0
  bandwidth 256
  no ip address
  encapsulation ppp
  no ip mroute-cache
  load-interval 30
  no fair-queue
  ppp multilink
  multilink-group 1
```

cRTP on MLP Connections

Compressed RTP (cRTP) can have a dramatic impact on the amount of bandwidth each voice call uses. Prior to Cisco IOS Release 12.0(7)T, cRTP was process switched. In fact, fast switching for cRTP was not available on the Catalyst 2600 and 3600 until a bug fix was implemented in Cisco IOS Release 12.0(7)T. In addition, some of the newer versions of Cisco IOS (specifically, Release 12.1(2.x)T) still use process switching for cRTP. Always read the release notes before attempting to use any specific feature.

The following example illustrates the commands used for this type of configuration:

```
interface Multilink1
  ip address 10.1.61.1 255.255.255.0
  ip tcp header-compression iphc-format
  no ip mroute-cache
  load-interval 30
  service-policy output QoS-Policy
  ppp multilink
  ppp multilink fragment-delay 10
  ppp multilink interleave
  multilink-group 1
  ip rtp header-compression iphc-format
```

LLQ for VoIP over MLP

Low-Latency Queuing (LLQ) is required to support voice over the WAN. When configuring LLQ for MLP-enabled interfaces, put the **service-policy output** in the multilink interface configuration. In the following example, two classes are defined: one for the VoIP media stream and one for the control traffic. Access to these classes, and therefore the queues they service, is done through access lists that match either Layer 3 ToS classification or source and destination IP addresses and ports. The access lists look slightly different for the control traffic at the central site because a Catalyst 6000 has already classified VoIP Control sessions with a DSCP value of 26 (AF31, which is backward compatible with IP Precedence 3).

All VoIP media traffic is placed into the Priority Queue (PQ), which is given 100 kbps of bandwidth. All Skinny Protocol control traffic is placed into a class-based queue and is given 10 kbps of bandwidth. All other traffic is queued using Weighted Fair Queuing.

The following example illustrates the commands used for this type of configuration:

```
class-map VoIP-RTP
  match access-group 100
class-map VoIP-Control
  match access-group 101
!
policy-map QoS-Policy-256k
  class VoIP-RTP
    priority 100
  class VoIP-Control
    bandwidth 8
  class class-default
    fair-queue
!
interface Multilink1
  ip address 10.1.61.1 255.255.255.0
  ip tcp header-compression iphc-format
  no ip mroute-cache
  load-interval 30
  service-policy output QoS-Policy
  ppp multilink
  ppp multilink fragment-delay 10
  ppp multilink interleave
  multilink-group 1
  ip rtp header-compression iphc-format
!
!      ToS VoIP Media Stream Classification: either IP Prec or DSCP
! This access-list is the same at the both the remote and
! central locations
access-list 100 permit ip any any precedence 5
access-list 100 permit ip any any dscp ef
!
! Skinny, H.323 and MGCP VoIP Control Traffic
! which has already been classified using the
! route-map in section 4.5.
access-list 101 permit ip any any precedence 3
access-list 101 permit ip any any dscp 26
```

Verifying Queuing, Fragmentation, and Interleaving on an MLP Connection

To verify the configuration settings, use the following commands (shown with their associated output):

```
1750# sh queue multilink1
Input queue: 1/75/0/0 (size/max/drops/flushes); Total output drops: 8288
Queueing strategy: weighted fair
Output queue: 63/1000/64/8288/1967(size/maxtotal/threshold/drops/interleaves)
  Conversations 1/3/256 (active/max active/max total)
  Reserved Conversations 1/1 (allocated/max allocated)

! All drops and interleaves are occurring on ToS=0 flows
(depth/weight/discards/tail drops/interleaves) 63/32384/8288/0/1967
Conversation 60, linktype: ip, length: 1008
source: 10.1.1.60.98, destination: 10.1.1.10.98, id: 0x0322, ttl: 63,
TOS: 0 prot: 17, source port 1024, destination port 7
```

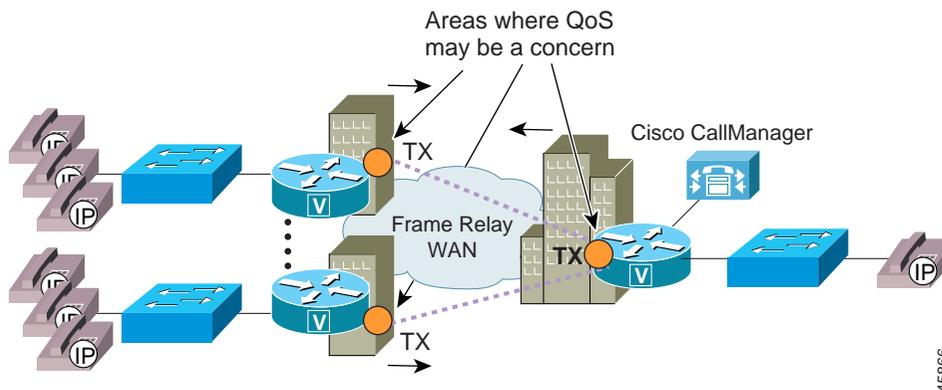
```
1750# sh policy interface multilink1
Multilink1
output : QoS-Policy-256k
Class VoIP-RTP
  Weighted Fair Queueing
    Strict Priority
    Output Queue: Conversation 264
      Bandwidth 100 (kbps)
      (pkts matched/bytes matched) 28100/5675882
      (pkts discards/bytes discards) 0/0
Class VoIP-Control
  Weighted Fair Queueing
    Output Queue: Conversation 265
      Bandwidth 8 (kbps) Max Threshold 64 (packets)
      (pkts matched/bytes matched) 204/10284
      (pkts discards/bytes discards/tail drops) 0/0/0
Class class-default
  Weighted Fair Queueing
    Flow Based Fair Queueing
    Maximum Number of Hashed Queues 256
```

Frame-Relay WAN

Frame-Relay networks are the most popular WANs in use today because of the low cost associated with them. However, because Frame Relay is a nonbroadcast technology that uses oversubscription to achieve costs savings, it is not always an easy platform on which to implement Cisco AVVID solutions. While this section outlines the basic requirements for successfully deploying Cisco AVVID solutions across a Frame-Relay WAN, extensive explanations of Frame Relay committed information rate (CIR), committed burst rate (Bc), excess burst rate (Be), and interval configurations are not covered here.

Figure 5-8 shows the general model for Frame-Relay WANs described in this guide.

Figure 5-8 General Model for a Frame-Relay WAN



When designing a Frame-Relay WAN for a Cisco AVVID network, keep the following recommendations in mind:

- Cisco IOS Release 12.1(2)T is the minimum recommended release for a Frame-Relay WAN.
- You must use traffic shaping with Frame-Relay WANs.
- Use Link Fragmentation and Interleaving (LFI) techniques on all virtual circuits with speeds below 768 kbps.

- Use Low-Latency Queuing (LLQ) with a Priority Queue (PQ) for VoIP bearer streams and a class-based queue for VoIP control sessions.
- Call admission control is required when the number of calls across the WAN can oversubscribe the allocated VoIP bandwidth.

The following sections explain the QoS issues for this type of configuration.

Traffic Shaping

Traffic shaping is required for Frame-Relay networks for three reasons:

- Oversubscription of sites is part of the nature of Frame-Relay networks.
- It is common for configurations to allow bursts that exceed the Committed Information Rate (CIR).
- The default interval for Cisco Frame-Relay devices can add unnecessary delay.

The following sections describe some of the aspects of traffic shaping for Frame-Relay networks.

Committed Information Rate

In most Frame-Relay networks, a central site uses a T1 link or something faster to terminate WAN connections from many remote offices. The central site sends data out at 1.536 Mbps, while a remote site may have only a 56-kbps circuit. In addition, there is typically a many-to-one ratio of remote offices to central hubs. It is quite possible for all the remote sites to send traffic at a rate that can overwhelm the T1 at the hub. Both of these scenarios can cause frame buffering in the provider network that induces delay, jitter, and drops. The only solution is to use traffic shaping at both the central and remote routers.

Committed Burst Rate

Another problem with Frame-Relay networks is the amount of data a node can transmit at any given time. A 56-kbps Permanent Virtual Circuit (PVC) can transmit a maximum of 56 kbits of traffic in 1 second. How this second is divided is called the *interval*. The amount of traffic a node can transmit during this interval is called the committed burst (Bc) rate. By default, all Cisco routers set Bc to CIR/8. The formula for calculating the interval is

$$\text{Interval} = \text{Bc}/\text{CIR}$$

For example, with a CIR of 56 kbps:

$$\text{Interval} = 7000 / 56,000 = 125 \text{ ms}$$

In the preceding example, after a router sends its allocated 7000 bits, it must wait 125 ms before sending its next traffic. While this is a good default value for data, it is a very bad choice for voice. By setting the Bc value to a much lower number, you can decrease the interval, which means the router will send traffic more frequently. An optimal configured value for Bc is 1000.

Excess Burst Rate

If the router does not have enough traffic to send all of its Bc (1000 bits, for example), it can "credit" its account and send more traffic during a later interval. The excess burst (Be) rate defines the maximum amount that can be credited to the router's traffic account. The problem with Be in Cisco AVVID networks is that this can create a potential for buffering delays within a Frame-Relay network because the receiving side can "pull" the traffic from a circuit only at the rate of Bc, not Bc + Be.

Minimum CIR

Cisco IOS defaults to a minimum CIR (**mincir**) value of CIR/2. Minimum CIR is the transmit value a Frame-Relay router will "rate down" to when BECNs are received.



Note

The maximum configured bandwidth in the Priority Queues (PQs) and class-based queues cannot exceed the minimum available amount of bandwidth on the WAN connection.

The following example shows a configuration for a remote site router connected to a 256-kbps Frame-Relay circuit:

```
interface Serial1
  no ip address
  encapsulation frame-relay
  load-interval 30
  frame-relay traffic-shaping
!
interface Serial1.71 point-to-point
  bandwidth 256
  ip address 10.1.71.1 255.255.255.0
  frame-relay interface-dlci 71
    class VoIP-256kbs
!
map-class frame-relay VoIP-256kbs
  frame-relay cir 256000
  frame-relay bc 1000
  frame-relay be 0
  frame-relay mincir 256000
  no frame-relay adaptive-shaping
  service-policy output QoS-Policy-256k
  frame-relay fragment 320
```

FRF.12 for LFI on Frame-Relay WANs

To enable Link Fragmentation and Interleaving (LFI) on Frame-Relay WANs, you must also use traffic shaping. Unlike MLP, the actual fragment size must be configured when using LFI on Frame Relay. In Frame-Relay networks, the fragmentation size is based on the Permanent Virtual Circuit (PVC), not the actual serialization rate (clocking speed) of the interface. This method is used because the Frame-Relay traffic shaping policy allows only the specified bit rate in the Committed Information Rate (CIR) to enter the interface transmit buffer. In other words, the rate of the PVC CIR is the clocking rate to reference when estimating fragmentation requirements in a frame-relay environment.

The following example illustrates the commands used for this type of configuration:

```
map-class frame-relay VoIP-256kbs
  frame-relay cir 256000
  frame-relay bc 1000
  frame-relay be 0
  frame-relay mincir 256000
  no frame-relay adaptive-shaping
  service-policy output QoS-Policy-256k
  frame-relay fragment 320
```

cRTP on Frame-Relay Connections

Compressed RTP (cRTP) can have a dramatic impact on the amount of bandwidth each voice call uses. While cRTP fast switching was enabled with Cisco IOS Release 12.0(7)T, some of the newer releases of Cisco IOS (specifically, Release 12.1(2.x)T) still use process switching for cRTP. Always read the release notes before attempting to use any specific feature.

The following example illustrates the commands used for this type of configuration:

```
interface Serial1
  no ip address
  encapsulation frame-relay
  load-interval 30
  frame-relay traffic-shaping
  ip rtp header-compression iphc-format
```

LLQ for VoIP over Frame Relay

Low-Latency Queuing (LLQ) is required to support voice over the WAN. When configuring LLQ for Frame-Relay interfaces, put the **service-policy output** in the **map-class frame-relay** configuration section. In the following example, two classes are defined: one for the VoIP media stream and one for the control traffic. Access to these classes, and therefore the queues they service, is done through access lists that match either Layer 3 ToS classification or source and destination IP addresses and ports. The access lists look slightly different for the control traffic at the central site because a Catalyst 6000 has already classified VoIP Control sessions with a DSCP value of 26 (AF31, which is backward compatible with IP Precedence 3).

All VoIP media traffic is placed into the Priority Queue (PQ), which is given 100 kbps of bandwidth. All Skinny Protocol control traffic is placed into a class-based queue and given 10 kbps of bandwidth. All other traffic is queued using Weighted Fair Queuing.

The following example illustrates the commands used for this type of configuration:

```
class-map VoIP-RTP
  match access-group 100
class-map VoIP-Control
  match access-group 101
!
policy-map QoS-Policy-256k
  class VoIP-RTP
    priority 100
  class VoIP-Control
    bandwidth 8
  class class-default
    fair-queue
!
interface Serial1
  no ip address
  encapsulation frame-relay
  load-interval 30
  frame-relay traffic-shaping
!
interface Serial1.71 point-to-point
  bandwidth 256
  ip address 10.1.71.1 255.255.255.0
  frame-relay interface-dlci 71
  class VoIP-256kbs
!
map-class frame-relay VoIP-256kbs
  frame-relay cir 256000
  frame-relay bc 1000
  frame-relay be 0
  frame-relay mincir 256000
  no frame-relay adaptive-shaping
  service-policy output QoS-Policy-256k
  frame-relay fragment 160
!
!       ToS VoIP Media Stream Classification: either IP Prec or DSCP
! This access-list is the same at the both the remote and
! central locations
access-list 100 permit ip any any precedence 5
access-list 100 permit ip any any dscp ef
!
! Skinny, H.323 and MGCP VoIP Control Traffic
! which has already been classified using the
! route-map in section 4.5.
access-list 101 permit ip any any precedence 3
access-list 101 permit ip any any dscp 26
```

Verifying Frame Relay Queuing, Fragmentation, and Interleaving

To verify the configuration settings, use the following commands (shown with their associated output):

```
3600# sh policy interface s 0/1.73
Remote Branch 3600
Serial0/1.73: DLCI 73 -

Service-policy output: QoS-Policy-256k (1117)

Class-map: VoIP-RTP (match-all) (1118/2)
  5008 packets, 964953 bytes
  30 second offered rate 0 bps, drop rate 0 bps
  Match: ip precedence 5 (1120)
  Weighted Fair Queueing
    Strict Priority
    Output Queue: Conversation 40
      Bandwidth 100 (kbps)
      (pkts matched/bytes matched) 4976/955161
      (pkts discards/bytes discards) 0/204

Class-map: VoIP-Control (match-all) (1122/3)
  53 packets, 3296 bytes
  30 second offered rate 0 bps, drop rate 0 bps
  Match: ip precedence 3 (1124)
  Weighted Fair Queueing
    Output Queue: Conversation 41
      Bandwidth 8 (kbps) Max Threshold 64 (packets)
      (pkts matched/bytes matched) 53/3296
      (pkts discards/bytes discards/tail drops) 0/0/0

Class-map: class-default (match-any) (1126/0)
  5329 packets, 985755 bytes
  30 second offered rate 0 bps, drop rate 0 bps
  Match: any (1128)
    5329 packets, 985755 bytes
    30 second rate 0 bps
  Weighted Fair Queueing
    Flow Based Fair Queueing
    Maximum Number of Hashed Queues 32
```

```
HQ_7200# sh frame-relay pvc int s6/0 73
Headquarters 7200
```

```
PVC Statistics for interface Serial6/0 (Frame Relay DTE)
```

```
DLCI = 73, DLCI USAGE = LOCAL, PVC STATUS = ACTIVE, INTERFACE = Serial6/0.73
```

```
input pkts 114          output pkts 103          in bytes 8537
out bytes 10633         dropped pkts 0           in FECN pkts 0
in BECN pkts 0         out FECN pkts 0         out BECN pkts 0
in DE pkts 0           out DE pkts 0
out bcast pkts 62      out bcast bytes 5203
pvc create time 00:04:22, last time pvc status changed 00:04:22
service policy QoS-Policy-256k
```

```
Service-policy output: QoS-Policy-256k (1099)
```

```
Class-map: VoIP-RTP (match-all) (1100/2)
  0 packets, 0 bytes
  30 second offered rate 0 bps, drop rate 0 bps
  Match: ip dscp 46 (1102)
  Weighted Fair Queueing
  Strict Priority
  Output Queue: Conversation 72
  Bandwidth 100 (kbps)
  (pkts matched/bytes matched) 0/0
  (pkts discards/bytes discards) 0/0
```

```
Class-map: VoIP-Control (match-all) (1104/3)
  25 packets, 3780 bytes
  30 second offered rate 0 bps, drop rate 0 bps
  Match: ip dscp 26 (1106)
  Weighted Fair Queueing
  Output Queue: Conversation 73
  Bandwidth 8 (kbps) Max Threshold 64 (packets)
  (pkts matched/bytes matched) 25/3780
  (pkts discards/bytes discards/tail drops) 0/0/0
```

```
Class-map: class-default (match-any) (1108/0)
  163 packets, 15708 bytes
  30 second offered rate 0 bps, drop rate 0 bps
  Match: any (1110)
  163 packets, 15708 bytes
  30 second rate 0 bps
  Weighted Fair Queueing
  Flow Based Fair Queueing
  Maximum Number of Hashed Queues 64
```

ATM WAN

```

Output queue size 0/max total 600/drops 0
fragment type end-to-end      fragment size 160
cir 768000    bc 7680         be 0          limit 960    interval 10
mincir 768000  byte increment 960    BECN response no
frags 125     bytes 10913     frags delayed 125   bytes delayed 10913

```

```

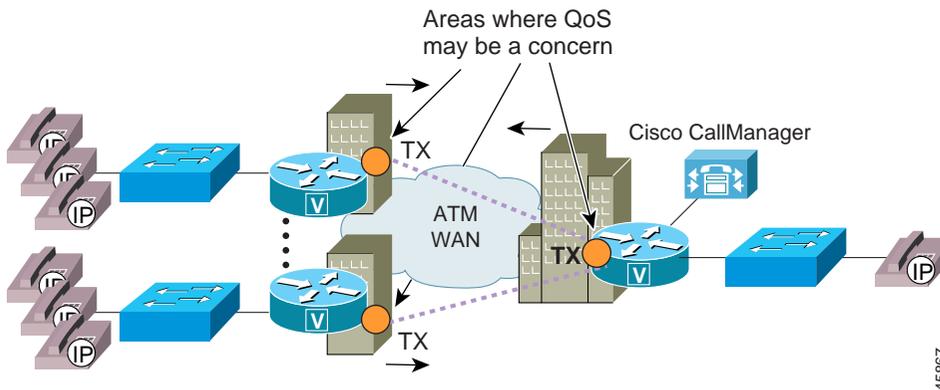
shaping inactive
traffic shaping drops 0

```

ATM WAN

Asynchronous Transfer Mode (ATM) is becoming a more common medium for WANs because many service providers have adopted this technology. Figure 5-9 shows the general model for ATM WANs described in this guide.

Figure 5-9 General Model for an ATM WAN



One of the difficulties with using ATM in WANs is that it was designed for high speeds, not low speeds. Many enterprises are attempting to deploy Cisco AVVID solutions over low-speed ATM connections. This generally results in complications because many of the Cisco IOS QoS tools are not currently supported on ATM interfaces, and many of the interface defaults are automatically configured for high-speed ATM circuits.

This is evident in the default sizing of ATM TX-ring buffers. For example, by default, the Cisco 7200 router OC-3 interface (the PA-A3) sets the TX-ring buffer to 8192 bytes. This is a correct setting for an OC-3, but, for a 256-kbps Permanent Virtual Circuit (PVC) configured on the interface, very large TX-ring buffer delays can occur. Because of this, the TX-ring has to be configured to a much lower value on a subinterface level. For example, the following configuration is for a remote site router connected to a 256-kbps ATM PVC:

```
interface ATM2/0
  no ip address
  no ip mroute-cache
  atm pvc 1 0 16 ilmi
  no atm ilmi-keepalive
!
interface ATM2/0.37 point-to-point
  pvc cisco37 0/37
    tx-ring-limit 7
    abr 256 256
    service-policy output QoS-Policy-256k
    protocol ppp Virtual-Template2
!
!
```

When designing an ATM WAN for a Cisco AVVID network, keep the following recommendations in mind:

- Cisco IOS Release 12.1(5)T for MLP over ATM is the minimum recommended release for an ATM WAN.
- For all ATM connections below DS-3 speeds, you must adjust the TX-ring buffer size.
- It is preferable to use two Permanent Virtual Circuits (PVCs) if the PVC speed is under 768 kbps.
- If using a single PVC that is under 768 kbps, use MLP over ATM for LFI.
- If using a single PVC, use LLQ with a Priority Queue (PQ) for VoIP bearer streams and a class-based queue for VoIP control sessions.
- Call admission control is required when the number of calls across the WAN can oversubscribe the allocated VoIP bandwidth.

The following sections explain the QoS issues for this type of configuration.

Two PVCs or LFI on Low-Speed ATM WANs

The best method of designing VoIP for ATM networks when using PVCs lower than 768 kbps is to use separate PVCs for voice and data. The following example illustrates this type of configuration:

```
interface ATM2/0.38 point-to-point
  bandwidth 256
  ip address 10.1.38.52 255.255.255.0
  pvc cisco38 0/38
    service-policy output Data-Policy-128k
    vbr-nrt 128 128
    encapsulation aal5snap
interface ATM2/0.39 point-to-point
  bandwidth 256
  ip address 10.1.39.52 255.255.255.0
  pvc cisco39 0/39
    tx-ring-limit 7
    service-policy output VoIP-Policy-128k
    vbr-nrt 128 128
    encapsulation aal5snap
```

If two PVCs are not an acceptable design alternative, the other option is to use the new MLP-over-ATM tools for link fragmentation and interleaving (LFI). Because ATM is a cell technology using a fixed payload size, there are no inherent LFI tools. A new standard, which uses MLP over ATM, is available in Cisco IOS Release 12.1(5)T. MLP over ATM provides a Layer 2 fragmentation and interleaving method for low-speed ATM links.

The ideal fragment size for MLP over ATM should allow the fragments to fit into an exact multiple of ATM cells. It is important to include MLP and ATM Adaptation Layer 5 (AAL5) overhead in all fragmentation calculations. The header for MLP over ATM is 10 bytes, and the AAL5 packet overhead is 8 bytes.

The fragment size for MLP over ATM can be calculated as follows:

$$\text{Fragment_Size} = (48 * \text{Number_of_Cells}) - 10 - 8$$

For example, if 7 cells per fragment is desirable, the fragment size should be

$$\text{Fragment_Size} = (48 * 7) - 10 - 8 = 318 \text{ bytes}$$

There are some interesting features for MLP over ATM, including the use of Virtual Template instead of Multilink interfaces. (Virtual-Template configurations will be replaced by Multilink interfaces in later releases of MLP over ATM because Multilink interfaces provide more scalability and greater integration into

existing MLP installations.) In addition, the configuration of PPP Challenge Handshake Authentication Protocol (CHAP) is required if remote sites want to communicate using MLP over ATM.

MLP over ATM requires the MLP bundle to classify the outgoing packets before they are sent to the ATM virtual circuit (VC). It also requires FIFO queuing to be used as the per-VC queuing strategy for the ATM VC. To use the advanced Low-Latency Queuing (LLQ) recommended for all VoIP WAN installations, attach the LLQ logic to the virtual template interface.

Only certain advanced ATM hardware supports per-VC traffic shaping (for example, ATM Deluxe PA on the Cisco 7200 router and OC-3 NM on the Cisco 3600 series). Because traffic shaping is a fundamental requirement of this design, MLP over ATM can be supported only on the platforms that support this ATM hardware. The following example illustrates this type of configuration:

```
interface ATM2/0
  no ip address
  no ip mroute-cache
  atm pvc 1 0 16 ilmi
  no atm ilmi-keepalive
!
interface ATM2/0.37 point-to-point
  pvc cisco37 0/37
    tx-ring-limit 7
    abr 256 256
    protocol ppp Virtual-Template2
!
!
interface Virtual-Template2
  bandwidth 254
  ip address 10.1.37.52 255.255.255.0
  service-policy output QoS-Policy-256k
  ppp authentication chap
  ppp chap hostname HQ_7200
  ppp chap password 7 05080F1C2243
  ppp multilink
  ppp multilink fragment-delay 10
  ppp multilink interleave
```

cRTP on ATM Connections

Compressed RTP (cRTP) is not currently supported on ATM interfaces.

LLQ for VoIP over ATM

Low-Latency Queuing (LLQ) is required to support voice over the ATM WAN when a single PVC is used. When configuring LLQ for ATM-enabled interfaces, place the **service-policy output** under the subinterface PVC configuration section. In the following example, two classes are defined: one for the VoIP media stream and one for the control traffic. Access to these classes, and therefore the queues they service, is done through access lists that match either Layer 3 ToS classification or source and destination IP addresses and ports. The access lists look slightly different for the control traffic at the central site because a Catalyst 6000 has already classified VoIP Control sessions with a DSCP value of 26 (AF31, which is backward compatible with IP Precedence 3).

All VoIP media traffic is placed into the Priority Queue (PQ), which is given 100 kbps of bandwidth. All Skinny Protocol control traffic is placed into a class-based queue and given 10 kbps of bandwidth. All other traffic is queued using Weighted Fair Queuing.

The following example illustrates this type of configuration:

```
class-map VoIP-RTP
  match access-group 100
class-map VoIP-Control
  match access-group 101
!
policy-map QoS-Policy-256k
  class VoIP-RTP
    priority 100
  class VoIP-Control
    bandwidth 8
  class class-default
    fair-queue
!
interface ATM2/0
  no ip address
  no ip mroute-cache
  atm pvc 1 0 16 ilmi
  no atm ilmi-keepalive
!
interface ATM2/0.37 point-to-point
  pvc cisco37 0/37
    tx-ring-limit 7
    abr 256 256
    protocol ppp Virtual-Template2
!
```

```
!
interface Virtual-Template2
  bandwidth 256
  ip address 10.1.37.52 255.255.255.0
  service-policy output QoS-Policy-256k
  ppp authentication chap
  ppp chap hostname HQ_7200
  ppp chap password 7 05080F1C2243
  ppp multilink
  ppp multilink fragment-delay 10
  ppp multilink interleave
!
!
!
!       ToS VoIP Media Stream Classification: either IP Prec or DSCP
! This access-list is the same at the both the remote and
! central locations
access-list 100 permit ip any any precedence 5
access-list 100 permit ip any any dscp ef
!
! Skinny, H.323 and MGCP VoIP Control Traffic
! which has already been classified using the
! route-map in Chapter 4.
access-list 101 permit ip any any precedence 3
access-list 101 permit ip any any dscp 26
```

Frame-Relay-to-ATM Interworking WAN

Many enterprises are deploying Cisco AVVID networks that use Frame Relay at the remote sites and ATM at the central location. The conversion is accomplished through ATM-to-Frame-Relay Service Interworking (FRF.8) in the carrier network.

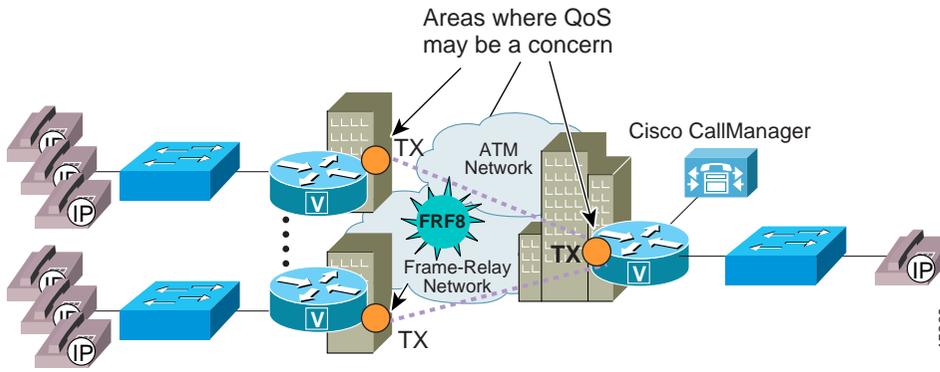


Note

When using MLP over ATM and Frame Relay for LFI, only Transparent Mode FRF.8 is supported.

Figure 5-10 shows the general model for a WAN using ATM at the central site and Frame Relay at the remote sites.

Figure 5-10 General Model of a WAN Combining ATM and Frame Relay



When designing a Frame-Relay-to-ATM Interworking WAN for a Cisco AVVID network, keep the following recommendations in mind:

- Cisco IOS Release 12.1(5)T for MLP over ATM and MLP over Frame Relay is the minimum recommended release for this configuration.
- FRF.8 Transparent Mode is the only support method for MLP over ATM and Frame-Relay Service Interworking.
- For all ATM connections below DS-3 speeds, you must adjust the TX-ring buffer size.
- Use two Permanent Virtual Circuits (PVCs) if the ATM and Frame-Relay PVC speed is under 768 kbps.
- If using a single PVC that is under 768 kbps, use MLP over ATM and Frame Relay for LFI.
- If using a single PVC, use LLQ with a Priority Queue (PQ) for VoIP bearer streams and a class-based queue for VoIP control sessions.
- Call admission control is required when the number of calls across the WAN can oversubscribe the allocated VoIP bandwidth.

The following sections explain the QoS issues for this type of configuration.

LFI on Low-Speed ATM-to-Frame-Relay Interworking WANs

FRF.12 cannot be used because currently no service provider supports FRF.12. In fact, no Cisco WAN switching gear supports FRF.12. Tunneling FRF.12 through the service provider network does not work because there is no FRF.12 standard on the ATM side. This is a problem because fragmentation is a requirement if any of the remote Frame-Relay sites use a circuit of 768 kbps or below. The best VoIP design for ATM networks when using PVCs lower than 768 kbps is to use separate PVCs for voice and data.

If two PVCs are not an acceptable design alternative, the other option is to use the new MLP over ATM and Frame-Relay tools for Link Fragmentation and Interleaving (LFI), available in Cisco IOS Release 12.1(5)T. MLP over ATM and Frame Relay provides an end-to-end Layer 2 fragmentation and interleaving method for low-speed ATM-to-Frame-Relay FRF.8 Service Interworking links.

FRF.8 Service Interworking is a Frame Relay Forum (FRF) standard for connecting Frame-Relay networks with ATM networks. Service Interworking provides a standards-based solution for service providers, enterprises, and end users. In Service Interworking translation mode, Frame-Relay PVCs are mapped to ATM PVCs without the need for symmetric topologies because the paths can terminate on the ATM side. FRF.8 supports two modes of operation of the Interworking Frame Relay (IWF) for upper-layer user protocol encapsulation, which differ in the following ways:

- Translation Mode — Maps between ATM and Frame-Relay encapsulation. It also supports interworking of routed or bridged protocols.
- Transparent Mode — Does not map encapsulations but sends them unaltered. This mode is used when translation is impractical because encapsulation methods do not conform to the supported standards for Service Interworking.



Note

MLP for LFI on ATM and Frame-Relay Service Interworking networks is supported only when Transparent Mode is used.

To make MLP over Frame Relay and MLP over ATM interworking possible, the interworking switch must be configured in Transparent Mode, and the end routers must be able to recognize headers for both MLP over Frame Relay and MLP over ATM. You can enable these options with the **frame-relay interface-dlci <dlci> ppp** and **protocol ppp** commands for Frame Relay and ATM, respectively.

When a frame is sent from the Frame-Relay side of an ATM-to-Frame-Relay Service Interworking connection, the following actions should occur to make interworking possible:

1. A packet is encapsulated in the MLP-over-Frame-Relay header by the sending router.
2. The carrier switch, in Transparent Mode, strips off the two-byte Frame-Relay data-link connection identifier (DLCI) field and sends the rest of the packet to its ATM interface.
3. The receiving router examines the header of the received packet. If the first two bytes of the received packet are 0x03cf, the router treats it as a legal MLP-over-ATM packet and sends it to the MLP layer for further processing.

When an ATM cell is sent from the ATM side of an ATM-to-Frame-Relay Service Interworking connection, the following actions should occur to make interworking possible:

1. A packet is encapsulated in the MLP-over-ATM header by the sending router.
2. The carrier switch, in Transparent Mode, prepends a two-byte Frame-Relay DLCI field to the received packet and sends the packet to its Frame-Relay interface.
3. The receiving router examines the header of the received packet. If the first four bytes after the two-byte data-link connection identifier (DLCI) field of the received packet are 0xfefe03cf, the router treats it as a legal MLP-over-Frame-Relay packet and sends it to the MLP layer for further processing.

A new ATM-to-Frame-Relay Service Interworking standard, FRF.8.1, supports MLP over ATM and Frame Relay-Service Interworking. However, it might be years before all switches are updated to this new standard.

The ideal fragment size for MLP over ATM should allow the fragments to fit into an exact multiple of ATM cells. It is important to include MLP and Adaptation Layer 5 (AAL5) overhead in all fragmentation calculations. The header for MLP over ATM is 10 bytes, and the AAL5 packet overhead is 8 bytes.

The fragment size for MLP over ATM can be calculated as follows:

$$\text{Fragment_Size} = (48 * \text{Number_of_Cells}) - 10 - 8$$

For example, if 7 cells per fragment is desirable, the fragment size should be

$$\text{Fragment_Size} = (48 * 7) - 10 - 8 = 318 \text{ bytes}$$

There are some interesting features for MLP over ATM, including the use of Virtual Template instead of Multilink interfaces. (Virtual-Template configurations will be replaced by Multilink interfaces in later releases of MLP over ATM because Multilink interfaces provide more scalability and greater integration into existing MLP installations.) In addition, the configuration of PPP Challenge Handshake Authentication Protocol (CHAP) is required if remote sites want to communicate using MLP over ATM.

MLP over ATM requires the MLP bundle to classify the outgoing packets before they are sent to the ATM virtual circuit (VC). It also requires FIFO queuing to be used as the per-VC queuing strategy for the ATM VC. To use the advanced Low-Latency Queuing (LLQ) recommended for all VoIP WAN installations, attach the LLQ logic to the virtual template interface.

Only certain advanced ATM hardware supports per-VC traffic shaping (for example, ATM Deluxe PA on the Cisco 7200 router and OC-3 NM on the Cisco 3600 series). Because traffic shaping is a fundamental requirement of this design, MLP over ATM can be supported only on the platforms that support this ATM hardware.

MLP over Frame Relay also has some interesting features, such as the fact that it relies on a Frame-Relay traffic shaping (FRTS) engine to control the flow of packets from the MLP bundle to the Frame-Relay virtual circuit (VC).

The following sections present example configurations for ATM at the central site and Frame Relay at the remote sites.

ATM Configuration at the Central Site

The following example illustrates an ATM configuration at the central site:

```
interface ATM2/0
    no ip address
    no ip mroute-cache
    atm pvc 1 0 16 ilmi
    no atm ilmi-keepalive
!
interface ATM2/0.37 point-to-point
    pvc cisco37 0/37
        tx-ring-limit 7
        abr 256 256
        protocol ppp Virtual-Template2
!
!
interface Virtual-Template2
    bandwidth 254
    ip address 10.1.37.52 255.255.255.0
    service-policy output QoS-Policy-256k
    ppp authentication chap
    ppp chap hostname HQ_7200
    ppp chap password 7 05080F1C2243
    ppp multilink
    ppp multilink fragment-delay 10
    ppp multilink interleave
```

Frame-Relay Configuration at Remote Sites

The following example illustrates a Frame-Relay configuration at the remote sites:

```
interface Serial6/0
  description T1 to Frame Relay switch
  no ip address
  encapsulation frame-relay
  load-interval 30
  no arp frame-relay
  frame-relay traffic-shaping
!
interface Serial6/0.73 point-to-point
  description 3640
  no arp frame-relay
  frame-relay interface-dlci 73 ppp Virtual-Template2
  class VoIP-256kbs
!
interface Virtual-Template2
  bandwidth 254
  ip address 10.1.37.51 255.255.255.0
  service-policy output QoS-Policy-256k
  ppp authentication chap
  ppp chap hostname R72HQ
  ppp chap password 7 05080F1C2243
  ppp multilink
  ppp multilink fragment-delay 10
  ppp multilink interleave
```

cRTP on ATM-to-Frame-Relay Connections

Compressed RTP (cRTP) is not currently supported on ATM interfaces.

LLQ for Voice over ATM and Frame Relay

The LLQ configurations for Frame Relay and ATM links when using Service Interworking are exactly the same as when using end-to-end MLP over ATM. For details, see the “LLQ for VoIP over ATM” section on page 5-34.

Summary

As described in this chapter, the following general guidelines and recommendations apply when configuring a WAN for use with Cisco AVVID solutions:

- Use Link Fragmentation and Interleaving (LFI) techniques on all WAN connections with speeds below 768 kbps.
- Use Low-Latency Queuing (LLQ) on all WAN VoIP connections.
- Traffic shaping is required for all Frame-Relay and ATM deployments.
- Use compressed RTP (cRTP) wherever possible.
- ATM WANs operating at speeds below 768 kbps must use MLP over ATM to reduce frame sizes. MLP over ATM is supported in Cisco IOS Release 12.1(5)T.
- Frame-Relay-to-ATM Interworking environments require MLP over ATM and Frame Relay to reduce frame sizes on low-speed connections. MLP over ATM and Frame Relay is supported in Cisco IOS Release 12.1(5)T.
- Call admission control is required when the number of calls across the WAN can oversubscribe the provisioned VoIP bandwidth.