



Cisco AVVID QoS Design Guide

Cisco CallManager Release 3.0(5)

Corporate Headquarters

Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA

<http://www.cisco.com>

Tel: 408 526-4000
800 553-NETS (6387)

Fax: 408 526-4100

Customer Order Number: DOC-7811549=
Text Part Number: 78-11549-01

THE SPECIFICATIONS AND INFORMATION REGARDING THE PRODUCTS IN THIS MANUAL ARE SUBJECT TO CHANGE WITHOUT NOTICE. ALL STATEMENTS, INFORMATION, AND RECOMMENDATIONS IN THIS MANUAL ARE BELIEVED TO BE ACCURATE BUT ARE PRESENTED WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. USERS MUST TAKE FULL RESPONSIBILITY FOR THEIR APPLICATION OF ANY PRODUCTS.

THE SOFTWARE LICENSE AND LIMITED WARRANTY FOR THE ACCOMPANYING PRODUCT ARE SET FORTH IN THE INFORMATION PACKET THAT SHIPPED WITH THE PRODUCT AND ARE INCORPORATED HEREIN BY THIS REFERENCE. IF YOU ARE UNABLE TO LOCATE THE SOFTWARE LICENSE OR LIMITED WARRANTY, CONTACT YOUR CISCO REPRESENTATIVE FOR A COPY.

The Cisco implementation of TCP header compression is an adaptation of a program developed by the University of California, Berkeley (UCB) as part of UCB's public domain version of the UNIX operating system. All rights reserved. Copyright © 1981, Regents of the University of California.

NOTWITHSTANDING ANY OTHER WARRANTY HEREIN, ALL DOCUMENT FILES AND SOFTWARE OF THESE SUPPLIERS ARE PROVIDED "AS IS" WITH ALL FAULTS. CISCO AND THE ABOVE-NAMED SUPPLIERS DISCLAIM ALL WARRANTIES, EXPRESSED OR IMPLIED, INCLUDING, WITHOUT LIMITATION, THOSE OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT OR ARISING FROM A COURSE OF DEALING, USAGE, OR TRADE PRACTICE.

IN NO EVENT SHALL CISCO OR ITS SUPPLIERS BE LIABLE FOR ANY INDIRECT, SPECIAL, CONSEQUENTIAL, OR INCIDENTAL DAMAGES, INCLUDING, WITHOUT LIMITATION, LOST PROFITS OR LOSS OR DAMAGE TO DATA ARISING OUT OF THE USE OR INABILITY TO USE THIS MANUAL, EVEN IF CISCO OR ITS SUPPLIERS HAVE BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

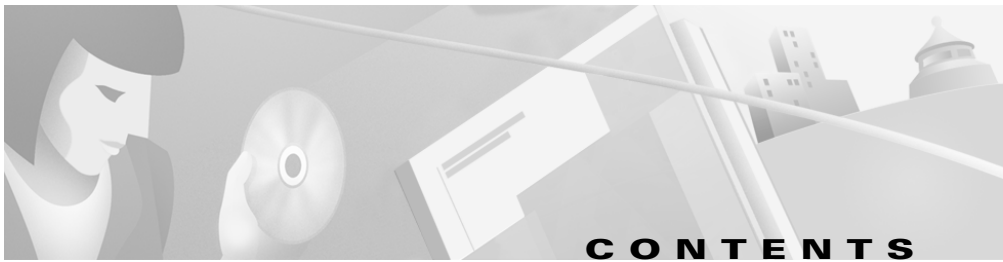
AtmDirector, Browse with Me, CCDA, CCDE, CCDP, CCIE, CCNA, CCNP, CCSI, CD-PAC, *CiscoLink*, the Cisco NetWorks logo, the Cisco Powered Network logo, Cisco Systems Networking Academy, the Cisco Systems Networking Academy logo, Fast Step, Follow Me Browsing, FormShare, FrameShare, GigaStack, IGX, Internet Quotient, IP/VC, iQ Breakthrough, iQ Expertise, iQ FastTrack, the iQ Logo, iQ Net Readiness Scorecard, MGX, the Networkers logo, *Packet*, PIX, RateMUX, ScriptShare, SlideCast, SMARTnet, TransPath, Voice LAN, Wavelength Router, WebViewer are trademarks of Cisco Systems, Inc.; Changing the Way We Work, Live, Play, and Learn, Empowering the Internet Generation, are service marks of Cisco Systems, Inc.; and Aironet, ASIST, BPX, Catalyst, Cisco, the Cisco Certified Internetwork Expert Logo, Cisco IOS, the Cisco IOS logo, Cisco Systems, Cisco Systems Capital, the Cisco Systems logo, Enterprise/Solver, EtherChannel, EtherSwitch, FastHub, IOS, IP/TV, LightStream, Post-Routing, Pre-Routing, Registrar, StrataView Plus, Stratm, SwitchProbe, TeleRouter, and VCO are registered trademarks of Cisco Systems, Inc. or its affiliates in the U.S. and certain other countries.

All other brands, names, or trademarks mentioned in this document or Web site are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (0011R)

Cisco AVVID QoS Design Guide

Copyright © 2000, 2001, Cisco Systems, Inc.

All rights reserved.



Preface xi

Purpose **xi**

Audience **xii**

Organization **xii**

Conventions **xiii**

Additional Information **xv**

Obtaining Documentation **xv**

 World Wide Web **xv**

 Documentation CD-ROM **xv**

 Ordering Documentation **xvi**

Obtaining Technical Assistance **xvi**

 Cisco Connection Online **xvi**

 Technical Assistance Center **xvii**

 Documentation Feedback **xviii**

CHAPTER 1

Overview 1-1

Why is QoS Needed? **1-1**

 Network Quality **1-2**

 Network Congestion **1-2**

 Delay and Jitter **1-2**

- QoS Tools **1-5**
 - Classification **1-5**
 - Queuing **1-7**
 - Network Provisioning **1-7**
- Summary **1-9**

CHAPTER 2**Connecting IP Phones 2-1**

- Using a Single Cable to Install an IP Phone **2-2**
 - Speed and Duplex Settings **2-3**
 - Catalyst 4000 and 6000 **2-4**
 - Catalyst 3500 XL and 2900 XL **2-5**
 - IP Addressing **2-5**
 - Catalyst 4000 and 6000 **2-6**
 - Catalyst 3500 XL and 2900 XL **2-7**
 - Classification and Queuing on the IP Phone **2-7**
 - Catalyst 6000 **2-9**
 - Catalyst 2948G, 2980G, and 4000 **2-10**
 - Catalyst 3500 XL and 2900 XL **2-10**
- Using Multiple Cables to Install an IP Phone **2-11**
 - Speed and Duplex **2-11**
 - IP Addressing **2-12**
 - Classification and Queuing on the IP Phone **2-12**
 - Catalyst 6000 **2-12**
 - Catalyst 4000 **2-13**
 - Catalyst 3500 XL and 2900 XL **2-14**

- Installing SoftPhone **2-15**
 - Speed and Duplex **2-15**
 - IP Addressing **2-15**
 - Classification and Queuing on the IP Phone **2-15**
- Using Separate Access Layer Switches for IP Phones **2-16**
 - Speed and Duplex **2-16**
 - IP Addressing **2-17**
 - Classification and Queuing on the IP Phone **2-17**
- Summary **2-18**

CHAPTER 3**Designing a Campus 3-1**

- Campus Switching Designs for Cisco AVVID **3-1**
 - Queue Scheduling **3-4**
 - Number of Queues **3-5**
- Marking Control and Management Traffic **3-6**
 - Skinny Protocol **3-8**
 - H.323 Protocol **3-9**
 - MGCP **3-10**
- Catalyst 6000 Access Layer **3-11**
 - Catalyst 6000 Port Scheduling and Queuing Schemes **3-13**
 - Receive Interface **3-13**
 - Transmit Interface **3-14**
 - Configuring QoS Parameters **3-15**
 - IP Phone Port Queuing **3-16**
 - Verifying IP Phone Access Port Configuration **3-17**
 - Uplink Interface to the Distribution Switch **3-21**
 - MLS and Catalyst QoS Configuration **3-21**

- Catalyst 6000 Transmit Queue Configuration **3-21**
- Catalyst 6000 CoS/ToS-to-DSCP Mapping Configuration **3-22**
- Verifying CoS/ToS-to-DSCP Mapping **3-22**
- Catalyst 4000 Access Layer **3-23**
 - Catalyst 4000 Port Scheduling and Queuing Schemes **3-23**
 - Receive Interface **3-23**
 - Transmit Interface **3-24**
 - Catalyst 4000 Switch-Wide QoS **3-25**
 - Verifying Catalyst 4000 Queue Admission Configuration **3-26**
 - IP Phone Port Queuing **3-26**
 - Uplink Interface to the Distribution Switch **3-26**
- Catalyst 3500 Access Layer **3-27**
 - Catalyst 3500 Port Scheduling and Queuing Schemes **3-28**
 - Receive Interface **3-28**
 - Transmit Interface 10/100 Ports **3-28**
 - Transmit Interface Gigabit Ethernet Ports **3-28**
 - IP Phone Port Queuing **3-30**
 - Uplink Interface to the Distribution Switch **3-30**
- Catalyst 6000 Distribution Layer **3-31**
 - Configuring Catalyst 6000 Distribution Layer VoIP Control Traffic Transmit Queue **3-32**
 - Catalyst 6000 Distribution Layer Configuration with a Catalyst 6000-PFC Access Layer **3-33**
 - Trust DSCP from the Layer 3 Access Switch **3-33**
 - Catalyst 6000 ToS-to-DSCP Mapping Configuration **3-34**

Catalyst 6000 Distribution Layer Configuration with an Access Switch Enabled for Layer 2 Only	3-34
Trust CoS from the Layer 2 Access Switch	3-35
Catalyst 6000 CoS-to-DSCP Mapping Configuration	3-35
Configuring Layer 3 Access Lists for VoIP Control Traffic Classification	3-36
Configuring the Connection to the Cisco 7200 WAN Router	3-37
Catalyst 6000 Distribution/Core Running Native IOS	3-38
Configuring QoS on the Native Cisco IOS Catalyst 6000	3-39
Configuring Transmit Queue Admission for VoIP Control Traffic	3-40
Catalyst 6000 Native Cisco IOS Distribution Layer Configuration with a Catalyst 6000-PFC Access Layer	3-40
Trust DSCP from the Layer 3 Access Switch	3-40
Native Cisco IOS ToS-to-DSCP Mapping Configuration for Layer 3 Access Switches	3-41
Catalyst 6000 Native Cisco IOS Distribution Layer Configuration with an Access Switch Enabled for Layer 2 Only	3-42
Trust CoS from the Layer 2 Access Switch	3-42
Native IOS CoS-to-DSCP Mapping Configuration for Layer 2 Access Switches	3-43
Configure the QoS Policies and Layer 3 Access Lists for VoIP Control Traffic Classification	3-43
Summary	3-46

CHAPTER 4

Building a Branch Office 4-1

Recommended Branch Office Designs **4-1**

Using 802.1Q for Trunking Separate Voice and Data Subnets at the Branch Office **4-4**

Catalyst 3600 Branch Office Router Using 802.1Q Trunking **4-5**

Catalyst 4000 Using 802.1Q Trunking **4-6**

Catalyst 3500 Using 802.1Q Trunking **4-6**

Using Secondary IP Addressing for Separate Voice and Data Subnets at the Branch Office **4-7**

Classifying VoIP Control Traffic at the Branch Office **4-7**

Using a Single Subnet at the Branch Office **4-9**

Cisco 1750 Single Subnet Configuration **4-9**

Catalyst 3500 Single Subnet Configuration **4-10**

Catalyst 2600 Single Subnet (no Trunking) Configuration **4-10**

Catalyst 4000 Single Subnet Configuration **4-11**

Summary **4-11**

CHAPTER 5

Implementing a Wide Area Network 5-1

WAN QoS Overview **5-1**

Classification **5-2**

Queuing **5-2**

Link Fragmentation and Interleaving **5-4**

Traffic Shaping **5-6**

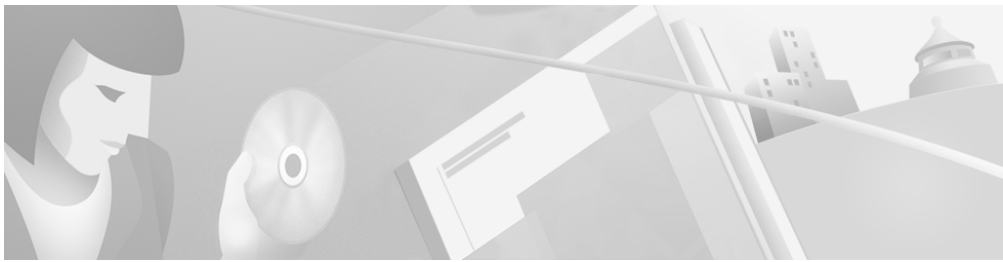
Network Provisioning **5-7**

Call Admission Control **5-10**

Miscellaneous WAN QoS Tools	5-11
VoIP Control Traffic	5-11
TX-Ring Sizing	5-12
Compressed Voice Codecs	5-14
Compressed RTP	5-14
Voice Activity Detection	5-15
Point-to-Point WAN	5-16
LFI on Point-to-Point WANs	5-17
cRTP on MLP Connections	5-18
LLQ for VoIP over MLP	5-18
Verifying Queuing, Fragmentation, and Interleaving on an MLP Connection	5-20
Frame-Relay WAN	5-21
Traffic Shaping	5-22
Committed Information Rate	5-22
Committed Burst Rate	5-23
Excess Burst Rate	5-23
Minimum CIR	5-24
FRF.12 for LFI on Frame-Relay WANs	5-25
cRTP on Frame-Relay Connections	5-26
LLQ for VoIP over Frame Relay	5-26
Verifying Frame Relay Queuing, Fragmentation, and Interleaving	5-28
ATM WAN	5-30
Two PVCs or LFI on Low-Speed ATM WANs	5-32
cRTP on ATM Connections	5-33
LLQ for VoIP over ATM	5-34

- Frame-Relay-to-ATM Interworking WAN **5-35**
 - LFI on Low-Speed ATM-to-Frame-Relay Interworking WANs **5-37**
 - ATM Configuration at the Central Site **5-40**
 - Frame-Relay Configuration at Remote Sites **5-41**
 - cRTP on ATM-to-Frame-Relay Connections **5-41**
 - LLQ for Voice over ATM and Frame Relay **5-41**
- Summary **5-42**

INDEX



Preface

This preface describes the purpose, intended audience, organization, and notational conventions for the *Cisco AVVID QoS Design Guide*.

Purpose

This document serves as an implementation guide for Voice over IP (VoIP) networks based on Cisco AVVID (Architecture for Voice, Video and Integrated Data). The goal of this document is to provide a blueprint for implementing the end-to-end Quality of Service (QoS) that is required for successful deployment of Cisco AVVID solutions in today's enterprise environment.

This document cannot examine all the possible QoS configurations available for all Cisco AVVID products. However, it does present configuration examples that are typical of the ones used in the majority of applications today. In particular, this document addresses QoS issues relating to

- High-speed campus designs
- Branch office solutions
- WAN implementations



Caution

The QoS design guidelines in this document are based on the best currently available knowledge about the functionality and operation of the Cisco AVVID components. The information in this document is subject to change without notice.

This document will be updated as the Cisco AVVID solution set grows with subsequent releases of Cisco CallManager and Cisco IOS.

Audience

This guide is intended for systems engineers and others responsible for designing VoIP networks based on Cisco AVVID solutions. This guide assumes that the reader has a basic knowledge of Cisco IOS, Cisco CatOS, Cisco AVVID products, and QoS theories in general.

Organization

The following table lists the chapters of this guide and the subjects they address:

Chapter	Title	Description
Chapter 1	Overview	Introduces QoS terms and concepts, and explains how they relate to VoIP networks.
Chapter 2	Connecting IP Phones	Describes several different methods for connecting IP phones to the VoIP network, and explains the QoS issues related to each method.
Chapter 3	Designing a Campus	Discusses the QoS issues involved with designing and implementing a VoIP network for the enterprise campus.
Chapter 4	Building a Branch Office	Discusses the QoS issue involved with connecting a branch office to the VoIP network.
Chapter 5	Implementing a Wide Area Network	Discusses the QoS issues involved with implementing VoIP over a Wide Area Network.

Conventions

This document uses the following conventions:

Convention	Description
boldface font	Commands and keywords are in boldface .
<i>italic font</i>	Arguments for which you supply values are in <i>italics</i> .
[]	Elements in square brackets are optional.
{ x y z }	Alternative keywords are grouped in braces and separated by vertical bars.
[x y z]	Optional alternative keywords are grouped in brackets and separated by vertical bars.
string	A nonquoted set of characters. Do not use quotation marks around the string or the string will include the quotation marks.
screen font	Terminal sessions and information the system displays are in <code>screen font</code> .
boldface screen font	Information you must enter is in boldface screen font .
<i>italic screen font</i>	Arguments for which you supply values are in <i>italic screen font</i> .
→	This pointer highlights an important line of text in an example.
^	The symbol ^ represents the key labeled Control—for example, the key combination ^D in a screen display means hold down the Control key while you press the D key.
< >	Nonprinting characters, such as passwords, are in angle brackets.

Notes use the following conventions:

**Note**

Means *reader take note*. Notes contain helpful suggestions or references to material not covered in the publication.

Timesavers use the following conventions:

**Timesaver**

Means *the described action saves time*. You can save time by performing the action described in the paragraph.

Tips use the following conventions:

**Tips**

Means *the information contains useful tips*.

Cautions use the following conventions:

**Caution**

Means *reader be careful*. In this situation, you might do something that could result in equipment damage or loss of data.

Warnings use the following conventions:

**Warning**

This warning symbol means danger. You are in a situation that could cause bodily injury. Before you work on any equipment, you must be aware of the hazards involved with electrical circuitry and familiar with standard practices for preventing accidents.

Additional Information

This section contains references to online documentation that provide additional information on subjects covered in this guide.

- Voice over IP and internetworking design:
 - http://www.cisco.com/univercd/cc/td/doc/product/voice/ip_tele/index.htm
 - <http://www.cisco.com/univercd/cc/td/doc/cisintwk/idg4/index.htm>
- High-availability design:
 - http://www.cisco.com/warp/partner/synchronicd/cc/sol/mkt/ent/ndsgn/highd_wp.htm
 - <http://www.zdnet.com/zdtag/whitepaper/campuslan.pdf>
- Glossary of terms and acronyms:
 - <http://www.cisco.com/univercd/cc/td/doc/cisintwk/ita/index.htm>
 - <http://www.cisco.com/univercd/cc/td/doc/product/voice/index.htm>

Obtaining Documentation

The following sections describe how to obtain this guide and other documents from Cisco.

World Wide Web

You can access the most current Cisco documentation on the World Wide Web at <http://www.cisco.com>, <http://www-china.cisco.com>, or <http://www-europe.cisco.com>.

Documentation CD-ROM

Cisco documentation and additional literature are available in a CD-ROM package, which ships with your product. The Documentation CD-ROM is updated monthly. Therefore, it is probably more current than printed documentation. The CD-ROM package is available as a single unit or as an annual subscription.

Ordering Documentation

Registered CCO users can order the Documentation CD-ROM and other Cisco Product documentation through our online Subscription Services at <http://www.cisco.com/cgi-bin/subcat/kaojump.cgi>.

Nonregistered CCO users can order documentation through a local account representative by calling Cisco's corporate headquarters (California, USA) at 408 526-4000 or, in North America, call 800 553-NETS (6387).

Obtaining Technical Assistance

Cisco provides Cisco Connection Online (CCO) as a starting point for all technical assistance. Warranty or maintenance contract customers can use the Technical Assistance Center. All customers can submit technical feedback on Cisco documentation using the web, e-mail, a self-addressed stamped response card included in many printed docs, or by sending mail to Cisco.

Cisco Connection Online

Cisco continues to revolutionize how business is done on the Internet. Cisco Connection Online is the foundation of a suite of interactive, networked services that provides immediate, open access to Cisco information and resources at anytime, from anywhere in the world. This highly integrated Internet application is a powerful, easy-to-use tool for doing business with Cisco.

CCO's broad range of features and services helps customers and partners to streamline business processes and improve productivity. Through CCO, you will find information about Cisco and our networking solutions, services, and programs. In addition, you can resolve technical issues with online support services, download and test software packages, and order Cisco learning materials and merchandise. Valuable online skill assessment, training, and certification programs are also available.

Customers and partners can self-register on CCO to obtain additional personalized information and services. Registered users may order products, check on the status of an order and view benefits specific to their relationships with Cisco.

You can access CCO in the following ways:

- WWW: www.cisco.com
- Telnet: cco.cisco.com
- Modem using standard connection rates and the following terminal settings: VT100 emulation; 8 data bits; no parity; and 1 stop bit.
 - From North America, call 408 526-8070
 - From Europe, call 33 1 64 46 40 82

You can e-mail questions about using CCO to cco-team@cisco.com.

Technical Assistance Center

The Cisco Technical Assistance Center (TAC) is available to warranty or maintenance contract customers who need technical assistance with a Cisco product that is under warranty or covered by a maintenance contract.

To display the TAC web site that includes links to technical support information and software upgrades and for requesting TAC support, use www.cisco.com/techsupport.

To contact by e-mail, use one of the following addresses:

Language	E-mail Address
English	tac@cisco.com
Hanzi (Chinese)	chinese-tac@cisco.com
Kanji (Japanese)	japan-tac@cisco.com
Hangul (Korean)	korea-tac@cisco.com
Spanish	tac@cisco.com
Thai	thai-tac@cisco.com

In North America, TAC can be reached at 800 553-2447 or 408 526-7209. For other telephone numbers and TAC e-mail addresses worldwide, consult the following web site:
<http://www.cisco.com/warp/public/687/Directory/DirTAC.shtml>.

Documentation Feedback

If you are reading Cisco product documentation on the World Wide Web, you can submit technical comments electronically. Click **Feedback** in the toolbar and select **Documentation**. After you complete the form, click **Submit** to send it to Cisco.

You can e-mail your comments to bug-doc@cisco.com.

To submit your comments by mail, for your convenience many documents contain a response card behind the front cover. Otherwise, you can mail your comments to the following address:

Cisco Systems, Inc.
Document Resource Connection
170 West Tasman Drive
San Jose, CA 95134-9883

We appreciate and value your comments.



Overview

This chapter presents an overview of the concepts and issues involved with maintaining Quality of Service (QoS) in an IP telephony network.

Why is QoS Needed?

Voice quality is directly affected by two major factors:

- Lost packets
- Delayed packets

Packet loss causes voice clipping and skips. The industry standard codec algorithms used in Cisco Digital Signal Processor (DSP) can correct for up to 30 ms of lost voice. Cisco Voice over IP (VoIP) technology uses 20-ms samples of voice payload per VoIP packet. Therefore, for the codec correction algorithms to be effective, only a single packet can be lost during any given time.

Packet delay can cause either voice quality degradation due to the end-to-end voice latency or packet loss if the delay is variable. If the end-to-end voice latency becomes too long (250 ms, for example), the conversation begins to sound like two parties talking on a CB radio. If the delay is variable, there is a risk of jitter buffer overruns at the receiving end. Eliminating drops and delays is even more imperative when including fax and modem traffic over IP networks. If packets are lost during fax or modem transmissions, the modems are forced to “retrain” to synchronize again. By examining the causes of packet loss and delay, we can gain an understanding of why Quality of Service (QoS) is needed in all areas of the enterprise network.

Network Quality

Voice packets can be dropped if the network quality is poor, if the network is congested, or if there is too much variable delay in the network. Poor network quality can lead to sessions frequently going out of service due to lost physical or logical connections. Because VoIP design and implementation is predicated on the assumption that the physical and logical network follows sound design methodologies and is extremely stable, network quality is not addressed in this guide.

Network Congestion

Network congestion can lead to both packet drops and variable packet delays. Voice packet drops from network congestion are usually caused by full transmit buffers on the egress interfaces somewhere in the network. As links or connections approach 100% utilization, the queues servicing those connections become full. When a queue is full, new packets attempting to enter the queue are discarded. This can occur on a campus Ethernet switch as easily as in the Frame Relay network of a service provider.

Because network congestion is typically sporadic, delays from congestion tend to be variable in nature. Egress interface queue wait times or large serialization delays cause variable delays of this type. Both of these factors are discussed in the next section, “Delay and Jitter”.

Delay and Jitter

Delay is the time it takes for a packet to reach the receiving endpoint after being transmitted from the sending endpoint. This time is termed the “end-to-end delay,” and it consists of two components: fixed network delay and variable network delay. Jitter is the delta, or difference, in the total end-to-end delay values of two voice packets in the voice flow.

Fixed network delay should be examined during the initial design of the VoIP network. The International Telecommunications Union (ITU) standard G.114 states that a one-way delay budget of 150 ms is acceptable for high voice quality. Research at Cisco has shown that there is a negligible difference in voice quality scores using networks built with 200-ms delay budgets. Examples of fixed

network delay include the propagation delay of signals between the sending and receiving endpoints, voice encoding delay, and the voice packetization time for various VoIP codecs. Propagation delay calculations work out to almost 0.0063 ms/km. The G.729A codec, for example, has a 25 ms encoding delay value (two 10 ms frames + 5 ms look-ahead) and an additional 20 ms of packetization delay.

Congested egress queues and serialization delays on network interfaces can cause variable packet delays. Without Priority or Low-Latency Queuing (LLQ), queuing delay times equal serialization delay times as link utilization approaches 100%. Serialization delay is a constant function of link speed and packet size. As shown in Table 1-1, the larger the packet and the slower the link clocking speed, the greater the serialization delay. While this is a known ratio, it can be considered variable because a larger data packet can enter the egress queue before a voice packet at any time. If the voice packet must wait for the data packet to serialize, the delay incurred by the voice packet is its own serialization delay plus the serialization delay of the data packet in front of it. Using Cisco Link Fragmentation and Interleave (LFI) techniques, discussed in Chapter 5, “Implementing a Wide Area Network,” serialization delay can be configured to be a constant delay value.

Table 1-1 *Serialization Delay as a Function of Link Speed and Packet Size*

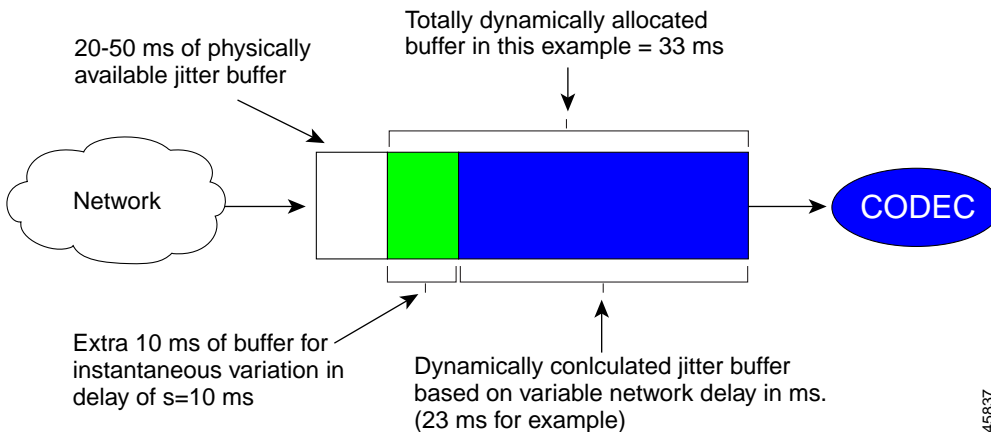
Link Speed	Packet Size					
	64 Bytes	128 Bytes	256 Bytes	512 Bytes	1024 Bytes	1500 Bytes
56 kbps	9 ms	18 ms	36 ms	72 ms	144 ms	214 ms
64 kbps	8 ms	16 ms	32 ms	64 ms	128 ms	187 ms
128 kbps	4 ms	8 ms	16 ms	32 ms	64 ms	93 ms
256 kbps	2 ms	4 ms	8 ms	16 ms	32 ms	46 ms
512 kbps	1 ms	2 ms	4 ms	8 ms	16 ms	23 ms
768 kbps	0.640 ms	1.28 ms	2.56 ms	5.12 ms	10.24 ms	15 ms

Because network congestion can be encountered at any time within a network, buffers can fill instantaneously. This instantaneous buffer utilization can lead to a difference in delay times between packets in the same voice stream. This difference, called jitter, is the variation between when a packet is expected to

arrive and when it actually is received. To compensate for these delay variations between voice packets in a conversation, VoIP endpoints use jitter buffers to turn the delay variations into a constant value so that voice can be played out smoothly.

Cisco VoIP endpoints use DSP algorithms that have an adaptive jitter buffer between 20 and 50 ms, as illustrated in Figure 1-1. The actual size of the buffer varies between 20 and 50 ms based on the expected voice packet network delay. These algorithms examine the timestamps in the Real-time Transport Protocol (RTP) header of the voice packets, calculate the expected delay, and adjust the jitter buffer size accordingly. When this adaptive jitter buffer is configured, a 10-ms portion of "extra" buffer is configured for variable packet delays. For example, if a stream of packets is entering the jitter buffer with RTP timestamps indicating 23 ms of encountered network jitter, the receiving VoIP jitter buffer is sized at a maximum of 33 ms. If a packet's jitter is greater than 10 ms above the expected 23-ms delay variation ($23 + 10 = 33$ ms of dynamically allocated adaptive jitter buffer space), the packet is dropped.

Figure 1-1 Adaptive Jitter Buffer



QoS Tools

Voice quality is only as good as the quality of the weakest network link. Packet loss, delay, and delay variation all contribute to degraded voice quality. In addition, because network congestion (or more accurately, instantaneous buffer congestion) can occur at any time in any portion of the network, network quality is an end-to-end design issue. The QoS tools discussed throughout this guide are a set of mechanisms to increase voice quality on data networks by decreasing dropped voice packets during times of network congestion and by minimizing both the fixed and variable delays encountered in a given voice connection.

These QoS tools can be separated into three categories:

- Classification
- Queuing
- Network provisioning

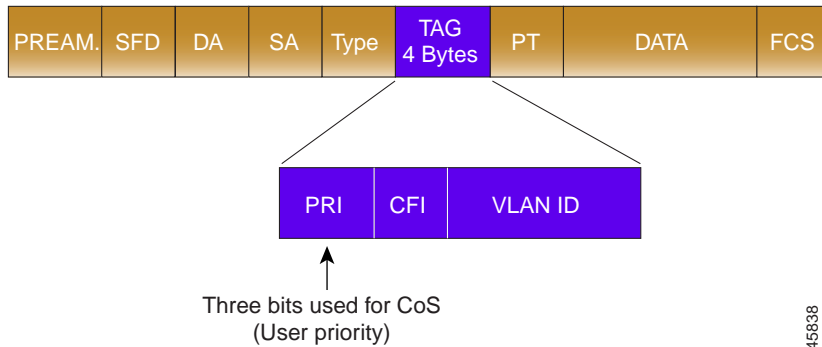
The following sections describe these categories.

Classification

Classification tools mark a packet or flow with a specific priority. This marking establishes a trust boundary that must be enforced.

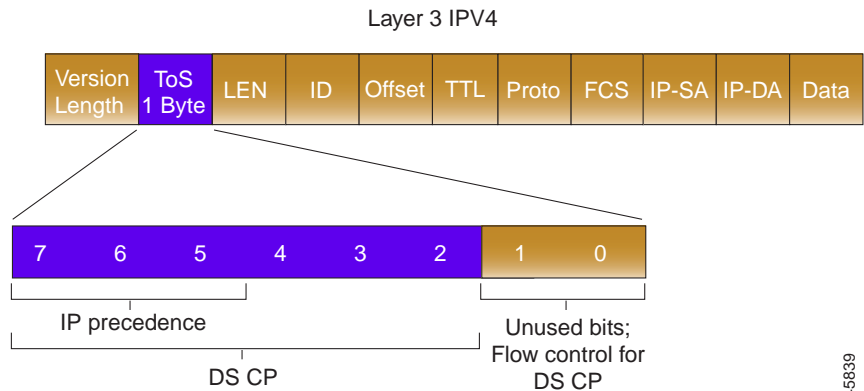
Classification should take place at the network edge, typically in the wiring closet or within the IP phones or voice endpoints themselves. Packets can be marked as important by using Layer 2 Class of Service (CoS) settings in the User Priority bits of the 802.1p portion of the 802.1Q header (see Figure 1-2) or the IP Precedence/Differentiated Services Code Point (DSCP) bits in the Type of Service (ToS) Byte of the IPv4 header (see Figure 1-3). All IP phone Real-time Transport Protocol (RTP) packets should be tagged with a CoS value of 5 for the Layer 2 802.1p settings and an IP Precedence value of 5 for Layer 3 settings. In addition, all Control packets should be tagged with a Layer 2 CoS value of 3 and a Layer 3 ToS of 3. Table 1-2 lists the CoS, IP Precedence, and DSCP settings for specifying packet priority.

Figure 1-2 Layer 2 Settings



45638

Figure 1-3 Layer 3 Settings



Standard IPv4: Three MSB called IP precedence
(DiffServ may use six D.S. bits plus two for flow control)

45639

Table 1-2 Packet Priority Classifications

Layer 2 Class of Service	IP Precedence	DSCP
CoS 0	Routine (IP precedence 0)	0-7
CoS 1	Priority (IP precedence 1)	8-15
CoS 2	Immediate (IP precedence 2)	16-23
CoS 3	Flash (IP precedence 3)	24-31

Table 1-2 Packet Priority Classifications (continued)

Layer 2 Class of Service	IP Precedence	DSCP
CoS 4	Flash-override (IP precedence 4)	32-39
CoS 5	Critical (IP precedence 5)	40-47
CoS 6	Internet (IP precedence 6)	48-55
CoS 7	Network (IP precedence 7)	56-63

The practice of using IP Precedence to mark traffic is a transitional step until all IP devices support DSCP. Ideally, in the future, all Cisco VoIP endpoints will use a DSCP value of Expedited Forwarding (EF) for the RTP voice bearer flows and a DSCP value of Assured Forwarding 31 (AF31) for VoIP Control traffic.

Chapter 2, “Connecting IP Phones,” discusses classification at length.

Queuing

Queuing tools assign a packet or flow to one of several queues, based on classification, for appropriate treatment in the network.

When data, voice, and video are placed in the same queue, packet loss and variable delay are much more likely to occur. By using multiple queues on egress interfaces and placing voice packets into a different queue than data packets, network behavior becomes much more predictable. Queuing is addressed in all sections of this guide because buffers can reach capacity in any portion of the network.

Addressing serialization delay is considered part of an overall queuing solution. Because serialization delay is a factor only on slow-speed links (links of 768 kbps or below), it is addressed in Chapter 5, “Implementing a Wide Area Network.”

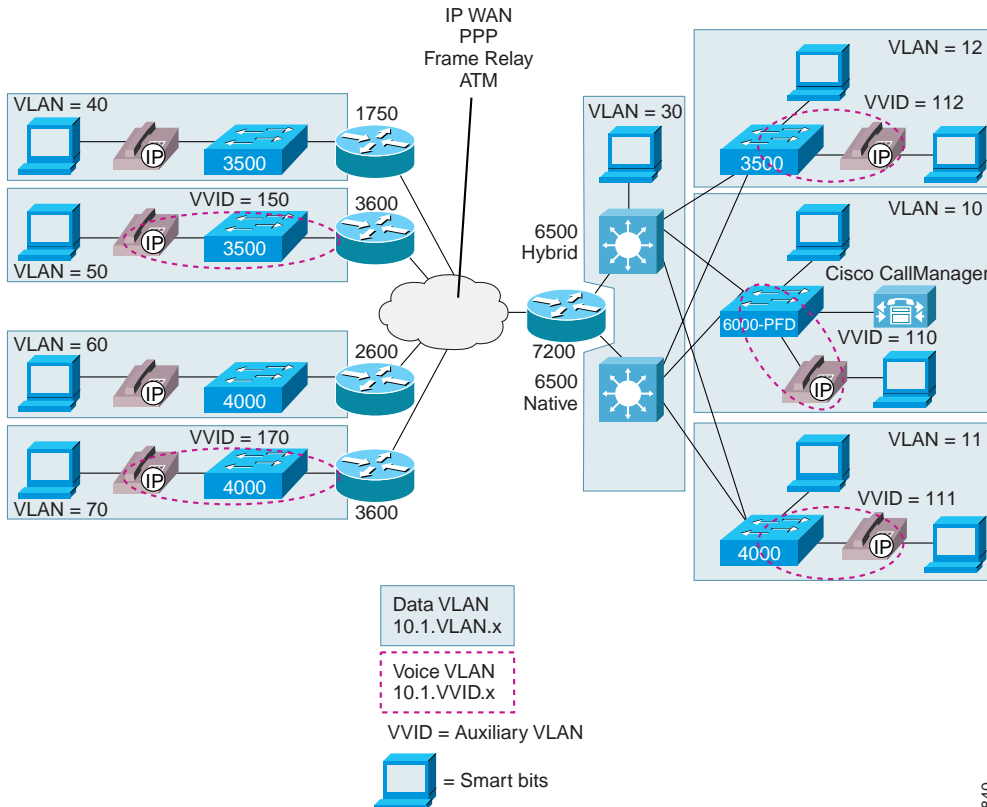
Network Provisioning

Network Provisioning tools accurately calculate the required bandwidth needed for voice conversations, all data traffic, any video applications, and necessary link management overhead such as routing protocols.

When calculating the required amount of bandwidth for running voice over a Wide Area Network, it is important to remember that all the application traffic (that is, voice, video, and data traffic), when added together, should equal no more than 75% of the provisioned bandwidth. The remaining 25% is used for overflow and administrative overhead, such as routing protocols. VoIP bandwidth calculations, Asynchronous Transfer Mode (ATM) cell overhead, and other details involved in network provisioning are discussed in Chapter 5, “Implementing a Wide Area Network.”

The Cisco QoS tools and example configurations in this guide are modeled on the network depicted in Figure 1-4.

Figure 1-4 General VoIP Network Model



45840

Summary

Quality of Service (QoS) for Voice over IP (VoIP) is guaranteed when proper VoIP network design is combined with the new Cisco Catalyst products, the latest Cisco IOS releases, and Cisco CallManager call admission control technologies. When building a Cisco AVVID network, you should adhere to the following core principles:

- Use 802.1Q/p connections for the IP phones and use the Auxiliary VLAN for voice.
- Classify voice RTP streams as EF or IP Precedence 5 and place them into a second queue (preferably a priority queue) on all network elements.
- Classify Voice Control traffic as AF31 or IP Precedence 3 and place it into a second queue on all network elements.
- Enable QoS within the campus if LAN buffers are reaching 100% utilization.
- Always provision the WAN properly, allowing 25% of the bandwidth for overhead, routing protocols, Layer 2 link information, and other miscellaneous traffic.
- Use Low Latency Queuing (LLQ) on all WAN interfaces.
- Use Link Fragmentation and Interleaving (LFI) techniques for all link speeds below 768 kbps.

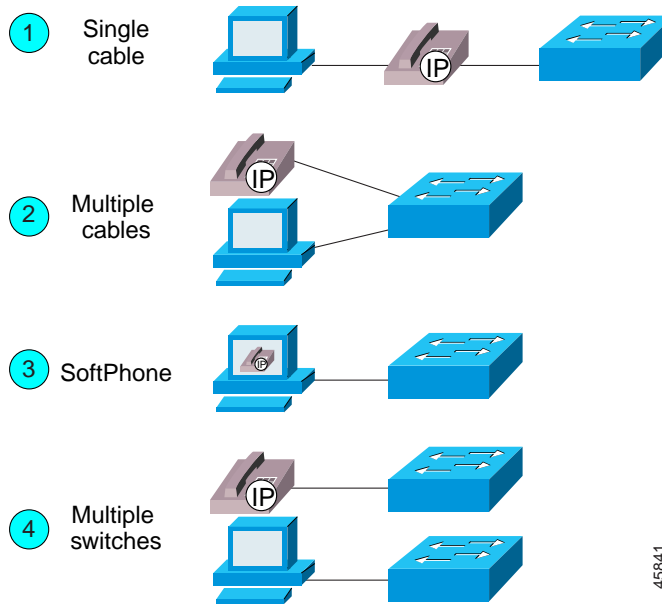
These principles are described more fully in subsequent chapters of this guide.



Connecting IP Phones

This chapter describes Quality of Service (QoS) issues relating to IP phones. As illustrated in Figure 2-1, there are essentially four ways to connect an IP phone to a campus network: using a single cable, using multiple cables, using the SoftPhone application running on a PC, or using separate switches for voice and data. Each of these connectivity methods has challenges for providing guaranteed voice quality. This chapter addresses those challenges, which can be summarized as follows:

- What speed and duplex settings should be used to connect an IP phone?
- What VLAN and IP addressing scheme should be used?
- How is classification and queuing handled for Voice over IP (VoIP) flows on an IP phone?

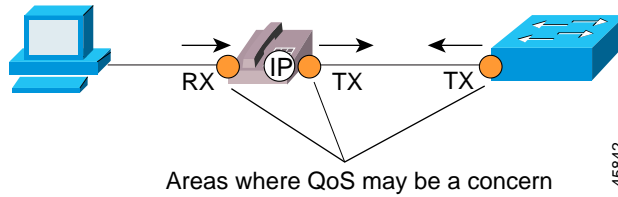
Figure 2-1 Ways to Connect IP Phones to the Network

45841

Using a Single Cable to Install an IP Phone

Most enterprises install IP phones on their Cisco AVVID network by using a single cable for both the phone and a PC. The reasons for this are ease of installation, savings on cabling infrastructure, and cost savings on wiring closet switch ports. With these cost savings comes requirements for additional switch features, particularly where QoS is concerned. Specifically, the requirements include correctly configuring the Ethernet link speed and duplex, Layer 2 Class of Service (CoS), and queuing on both the IP phone and the wiring closet Ethernet switch.

Figure 2-2 QoS Considerations When Using a Single-Cable Connection



Speed and Duplex Settings

The 10/100 Ethernet ports of the Cisco IP Phones support Auto-Negotiation for configuring speed and duplex. This is not user configurable. If the Network Interface Card (NIC) in the PC is also using Auto-Negotiation but the Ethernet switch port is configured for 10BaseT half-duplex, as many wiring closet switches are, a link speed mismatch could potentially lead to interface buffer overflow situations. While this half-duplex connection between the Cisco IP Phone and the switch should not normally be problematic, buffer congestion can arise through 100BaseT full-duplex to 10BaseT half-duplex aggregation. During periods of intense traffic (such as an extremely high-speed video stream), the half-duplex nature of the connection can lead to packet loss from deferred packets due to excessive collisions on the segment. Both the switch and the Cisco IP Phone, which uses a priority queue for voice, will always send voice traffic first. However, the high-speed video stream will also be sending as many packets as possible. When either the switch or the phone attempts to send the voice traffic (dependant, of course, on which direction the video flow is going), it can encounter collisions when attempting to transmit, thus resulting in deferred voice packets.

This extreme traffic loss example, while rare within a normal enterprise environment, is reproducible in the lab using SmartBits to simulate Motion Picture Experts Group (MPEG) video streams. This situation is best addressed by setting the switch port to Auto-Negotiation for all Cisco IP Phone connection options. If the port is statically set to 100BaseT full-duplex, the Cisco IP Phone automatically sets its port to 100BaseT half-duplex, resulting in a duplex mismatch. For details on why this duplex mismatch occurs, see

<http://www.cisco.com/warp/customer/473/3.html>

The use of Auto-Negotiation is the recommended configuration for IP phone connections because any user can modify his NIC configuration and enable 100BaseT full-duplex. In addition, using Auto-Negotiation enables 100BaseT full-duplex port speeds, which creates the infrastructure needed to support high-speed video applications.

Also, by using the CatOS PortFast mechanism, you can configure the phone access port to move into a forwarding state immediately, thereby decreasing IP phone boot time. To perform this configuration, use the **set port host** command on the Catalyst 4000 and 6000 or the **spanning-tree portfast** command on the 2900 XL and 3500 XL, which turns off Dynamic Trunking Protocol (DTP) and Port Aggregation Protocol (PAgP) and enables PortFast. For more details on this configuration, see the “Catalyst 4000 and 6000” and “Catalyst 3500 XL and 2900 XL” sections below.

**Note**

Phone boot times should not normally be a problem because the phone stays powered and connected at all times.

Catalyst 4000 and 6000

On the Catalyst 4000, 2948G, 2980G, and 6000 line of Ethernet switches, use the following commands to configure the ports properly for IP phones:

```
cat6k-access> (enable) set port inlinepower 5/1-48 auto
cat6k-access> (enable) set port speed 5/1-48 auto
cat6k-access> (enable) set port host 5/1-48
```

**Note**

Inline power is available only on power-enabled Ethernet line cards and is enabled by default.

Catalyst 3500 XL and 2900 XL

On the Catalyst 3500 XL and 2900 XL switches, use the following commands to configure the ports properly for IP phones:

```
interface FastEthernet0/1
  power inline auto
  speed auto
  spanning-tree portfast
```

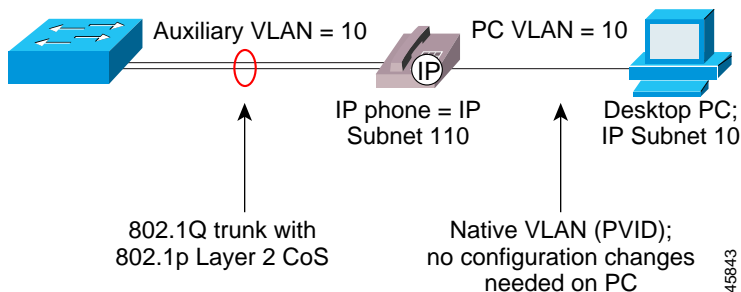
IP Addressing

After you have configured the speed and duplex settings for the IP phones, you need to consider IP addressing issues. There are three IP addressing options for the phones:

- Create a new subnet and use it for IP phones in a different IP address space (registered or RFC 1918 address space). Figure 2-3 illustrates this approach.
- Provide an IP address in the same subnet as the existing data device (PC or workstation).
- Start a new subnet in the existing IP address space. This might require redoing some or all of the IP addressing plan for the organization.

All of these options can be implemented using either Dynamic Host Configuration Protocol (DHCP) or static IP address configuration. Adding IP phones can potentially double your need for IP address space. While this may not be an issue in some enterprises, others may not have the available address space in particular subnets or even throughout the enterprise. These IP address space concerns, combined with the requirement of separation between the voice and data networks for administrative and QoS reasons, lead to the recommendation of creating a new subnet for the IP Phones.

Figure 2-3 Using a Separate Subnet for Voice

**Note**

Using a separate subnet, and potentially a separate IP address space, may not be an option for some small branch offices. Single address space configurations for connecting both IP phones and data devices is addressed in Chapter 4, “Building a Branch Office.”

Catalyst 4000 and 6000

The **set port auxiliaryvlan** command is a new CatOS command for creating IP phone 802.1Q access trunks in the Catalyst 2948G, 2980G, 4000, and 6000 family of switches. The following example uses this command in the Catalyst 6000:

```
cat6k-access> (enable) set vlan 10 name 10.1.10.0_data
cat6k-access> (enable) set vlan 110 name 10.1.110.0_voice
cat6k-access> (enable) set vlan 10 5/1-48
cat6k-access> (enable) set port auxiliaryvlan 5/1-48 110
```

The following example uses this command in the Catalyst 4000:

```
cat4k> (enable) set vlan 11 name 10.1.11.0_data
cat4k> (enable) set vlan 111 name 10.1.111.0_voice
cat4k> (enable) set vlan 11 2/1-48
cat4k> (enable) set port auxiliaryvlan 2/1-48 111
```

Catalyst 3500 XL and 2900 XL

To create the IP phone 802.1Q access trunks in the Catalyst 3500 XL and 2900 XL series, use the following commands:

```
interface FastEthernet0/1
  switchport trunk encapsulation dot1q
  switchport trunk native vlan 12
  switchport mode trunk
  switchport voice vlan 112
  spanning-tree portfast
vlan database
  vlan 112
```

**Note**

For ease of troubleshooting, the VLAN can be configured to match the subnet address.

Classification and Queuing on the IP Phone

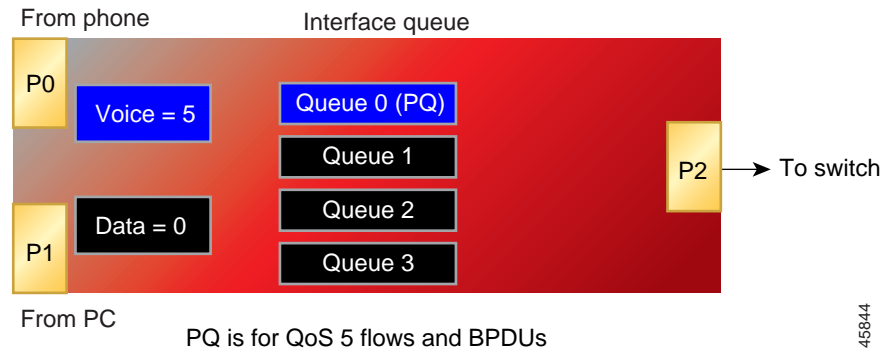
Classifying, or marking, traffic as close to the edge of the network as possible has always been an integral part of the Cisco network design architecture. When connected by a single cable, the IP phone is the edge of the managed network. As such, the IP phone can and should classify traffic flows.

Three User Priority bits in the 802.1p portion of the 802.1Q header are used for signaling Layer 2 CoS information (see Figure 1-2). By default, all VoIP Real-time Transport Protocol (RTP) bearer flows from the IP phone are set to a Layer 2 CoS value of 5 and a Layer 3 IP Precedence value of 5. Using IP Precedence is a transitional step, and all Cisco VoIP devices will eventually migrate to Differentiated Services Code Point (DSCP) for Layer 3 classification. At that time, Cisco VoIP endpoints (using DSCP instead of IP Precedence) will use a DSCP value of 46, or Expedited Forwarding (EF). These CoS and Type of Service (ToS) values are significant when examining how classification and queuing works both within an IP phone and in an enterprise network.

At the heart of a Cisco IP Phone is a three-port 10/100 switch. One port, P0, is an internal port used for connecting the actual voice electronics in the phone. Port P1 is used to connect a daisy-chained PC, and Port P2 is used to uplink to the wiring closet Ethernet switch. Each port has four queues with a single threshold (4QIT configuration). One of these queues, Queue 0, is a high-priority queue for all

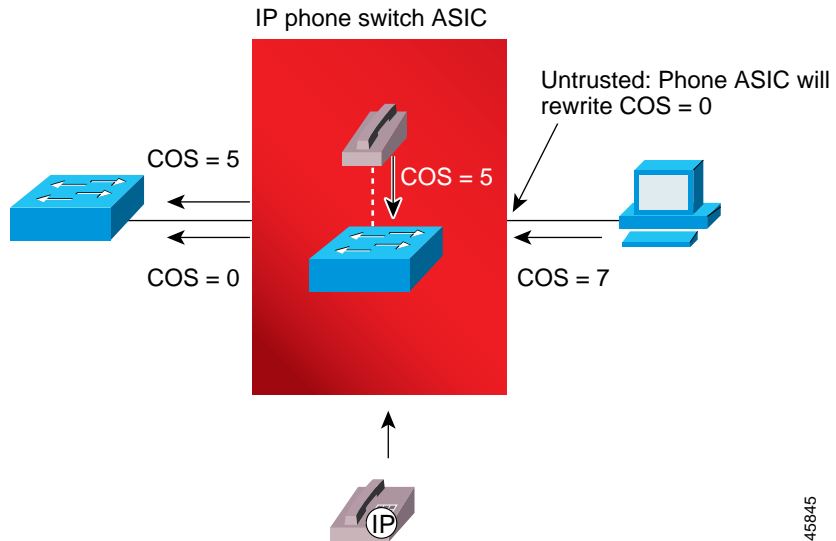
Bridge Protocol Data Unit (BPDU) and CoS=5 traffic. These queues are all serviced in a round-robin fashion with a timer used on the high-priority queue. If this timer expires while the queue scheduler is servicing the other queues, the scheduler automatically moves back to the high-priority queue and empties its buffer, ensuring voice quality. Figure 2-4 shows the queuing scheme for an IP phone.

Figure 2-4 Queuing for an IP Phone



Because the high-priority queue for the IP phone is accessible to any Layer 2 CoS=5 traffic, it is critical to make sure that the PC connected to the access port of the IP phone is not classifying traffic also. The recommended method for ensuring this is to extend the trust boundary of the Ethernet switch to the IP phone and not beyond, as illustrated in Figure 2-5.

Figure 2-5 Trust Boundaries for IP Phone



45845

Catalyst 6000

On Catalyst 6000 switches, trust boundary extension is done using the **set port qos trust-ext** command in Cisco CatOS Release 5.5. This command instructs the IP phone to mark traffic from the attached PC as CoS=0. Once the phone is configured to manipulate the CoS value, you must also configure the port to accept the CoS of the IP phone. On current Catalyst 10/100 Ethernet line cards, this requires a combination of configuration steps. First, the actual port has to be instructed to trust Layer 2 Ethernet CoS values from the IP phone. You can do this by using the **set port qos trust** commands. To overcome a configuration limitation on the line card ASIC, you must configure an additional Access Control List (ACL) to trust the IP phones. The best way to accomplish this is by

configuring an ACL to trust all CoS classification on Ethernet ports in the Auxiliary VLAN. The commands for establishing a classification and trust boundary at the IP phone are as follows:

```
cat6k-access> (enable) set port qos 5/1-48 trust-ext untrusted
cat6k-access> (enable) set port qos 5/1-48 trust trust-cos
cat6k-access> (enable) set qos acl ip ACL_IP-PHONES trust-cos ip any any
cat6k-access> (enable) set port qos 5/1-48 vlan-based
cat6k-access> (enable) commit qos acl all
cat6k-access> (enable) set qos acl map ACL_IP-PHONES 110
```

**Note**

The additional ACL configuration will not be required with the next generation of Catalyst 6000 10/100 line cards.

Catalyst 2948G, 2980G, and 4000

At the time of writing of this document, the Catalyst 2948G, 2980G, and 4000 do not currently offer the **set port qos <mod/port> trust trust-ext** commands. Therefore, these switches must rely on the default configuration of the IP phone, which uses CoS=5 for all VoIP streams and reclassifies CoS on all PC traffic to 0.

Catalyst 3500 XL and 2900 XL

When connecting IP phones to Catalyst 3500 XL and 2900 XL switches using the single-cable model, these switches require the same functionality as the Catalyst 6000 switches. To configure the IP phone not to trust the CoS settings from the PC, use the following commands:

```
interface FastEthernet0/1
    switchport priority extend cos 0
```

Using Multiple Cables to Install an IP Phone

You might want to use multiple cables to connect the IP phones if any of the following conditions apply to your Cisco IP Telephony network:

- You are connecting IP phones that do not have a second Ethernet port for attaching a PC.
- You want to create a physical separation between the voice and data networks.
- You want to provide in-line power easily to the IP phones without having to upgrade the data infrastructure.
- You want to limit the number of switches that need UPS power.
- You want to limit the amount of CatOS upgrades needed in the network.
- You want to limit the Spanning Tree configuration in the wiring closet switches.

Speed and Duplex

Because there is no PC behind the IP phone when you use multiple cables, port speed and duplex settings are not as critical as with a single cable. While it is safe to use the same configuration as with a single-cable connection (in case a PC is plugged into the second Ethernet port on the phone), this configuration is not required.

IP Addressing

The recommended configuration for using multiple cables to connect IP phones to the Cisco AVVID network is to use a separate IP subnet and separate VLANs for IP telephony.

**Note**

Using a separate subnet, and possibly a separate IP address space, may not be an option for some small branch offices due to the IP routing configuration. If the IP routing can handle an additional subnet at the remote branch, you can use Cisco Network Registrar and secondary addressing. Chapter 4, “Building a Branch Office,” discusses single address space configurations for connecting both IP phones and data devices.

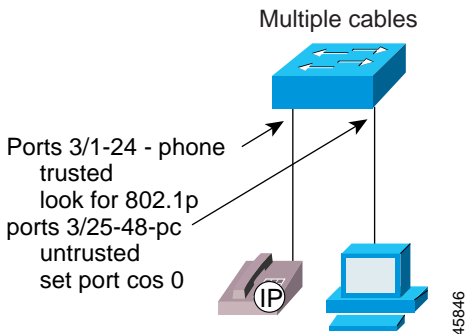
Classification and Queuing on the IP Phone

Since the IP phone and any data PCs are on separate physical cables, queuing on the IP phone is not required. However, since the IP phone is still a managed device, classification should still take place on the phone or ingress Access switch port. This classification for VoIP packets can be handled in a variety of ways, depending upon which hardware is used in the wiring closet switch. The following sections describe several different scenarios for the various types of switches.

Catalyst 6000

In the example shown in Figure 2-6, a Catalyst 6000 is used as a wiring closet switch. Ports 3/1-24 connect to IP phones, and ports 3/25-48 connect to data-only PCs. Because this is a tightly managed environment, all Layer 2 CoS settings are enforced on the Catalyst 6000.

Figure 2-6 Using Multiple Cables to Connect IP Phones to a Catalyst 6000



The commands for implementing this configuration are as follows:

```

cat6k-access> (enable) set port inlinepower 6/1-24 auto
cat6k-access> (enable) set port inlinepower 6/25-48 off
cat6k-access> (enable) set vlan 110 6/1-24
cat6k-access> (enable) set vlan 10 6/25-48
cat6k-access> (enable) set port auxiliaryvlan 6/1-24 dot1p
cat6k-access> (enable) set port host 6/1-24
cat6k-access> (enable) set port qos 6/1-24 trust-ext untrusted
cat6k-access> (enable) set port qos 6/1-24 trusttrust-cos
cat6k-access> (enable) set qos acl ip ACL_IP-PHONES trust-cos ip any any
cat6k-access> (enable) set port qos 6/1-24 vlan-based
cat6k-access> (enable) commit qos acl all
cat6k-access> (enable) set qos acl map ACL_IP-PHONES 110

```

Catalyst 4000

Currently there is no **dot1p** extension to the **auxiliaryvlan** command on the Catalyst 2948G, 2980G, and 4000 switches. To use the 802.1p classification of the IP phone for switch QoS, you must configure the Auxiliary VLAN with the same value as the port VLAN ID. This enables the IP phone to mark packets with the proper CoS settings.

The commands for implementing this configuration are

```

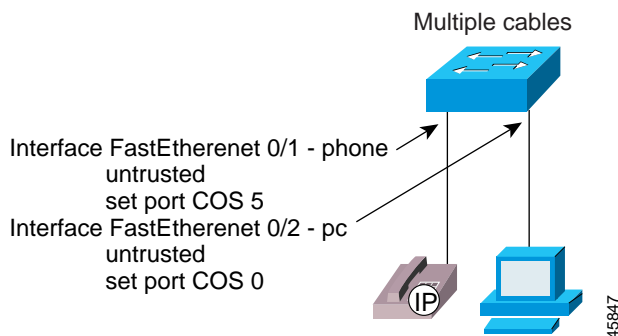
cat4k> (enable) set vlan 11 2/25-48
cat4k> (enable) set vlan 111 2/1-24
cat4k> (enable) set port host 2/1-48
cat4k> (enable) set port auxiliaryvlan 2/1-24 111

```

Catalyst 3500 XL and 2900 XL

Another option when configuring trust is to set it at the port level. On the Catalyst 3500 and 2900 XL series switches, either 802.1p or port-based CoS settings can be used for classifying traffic. A port-based configuration would look similar to the one in Figure 2-7.

Figure 2-7 Using Multiple Cables to Connect IP Phones to a Catalyst 3500 XL or 2900 XL



The commands for implementing this configuration are

```
interface FastEthernet0/1
  description IP Phone port
  spanning-tree portfast
  switchport mode access
  switchport access vlan 112
interface FastEthernet0/2
  description Data-only PC port
  switchport mode access
  switchport access vlan 12
```

Installing SoftPhone

Some companies deploy IP telephony using the Cisco SoftPhone application. In addition, many companies wish to evaluate the viability of using PC-based VoIP applications. Many PC network interface cards (NICs) do not currently set 802.1P CoS bits for Layer 2 classification. Even if the PCs did set Layer 2 frame markings, the vast majority of network administrators would resist "trusting" a user's PC. Because of these factors, Cisco SoftPhone currently classifies voice packets only at the Layer 3 IP header. In fact, all voice bearer packets originating from the Cisco SoftPhone application are marked with an IP Precedence value of 5. Of course, this marking requires a wiring closet Ethernet switch that is Layer 3 enabled, with multiple queues, to correctly queue these voice packets. Currently, this limits Cisco SoftPhone designs to PCs connected to Catalyst 6000 switches with a Policy Feature Card (PFC) installed.

Speed and Duplex

For use with SoftPhone, all wiring closet switch access ports should be set to full-duplex 100BaseT. Because PC CPUs are so fast, data has the potential to overrun voice on a half-duplex 10BaseT connection. See the "Using a Single Cable to Install an IP Phone" section on page 2-2 for details on setting speed and duplex on Catalyst switches.

IP Addressing

IP addressing is not an issue in this case because the SoftPhone application runs on a PC.

Classification and Queuing on the IP Phone

Cisco SoftPhone is a PC application, and it currently marks voice traffic only at the Layer 3 IP header. This requires the access switch to be Layer 3 aware because of the need to prioritize voice traffic before the first uplink to the distribution layer. This limits the choice of wiring closet switches with multiple queues to the Catalyst 6000 with a PFC installed.

**Note**

The switch must be configured to trust IP Precedence values from the Cisco SoftPhone application on the PC.

In the following example, a Catalyst 6000 is used as a wiring closet switch. The supervisor engine of the Catalyst 6000 has a PFC daughter card installed, which provides Layer 3/4 QoS intelligence. The access port connected to the PC is configured to trust all IP Precedence values originating from the PC. An ACL, `ACL_SOFTPHONE`, is also added as a workaround for the current configuration limitations on the Catalyst 6000 10/100 line cards. The commands for this configuration are

```
cat6k-access> (enable) set qos enable
cat6k-access> (enable) set port qos 7/1-48 port-based
cat6k-access> (enable) set port qos 7/1-48 trust trust-ipprec
cat6k-access> (enable) set qos acl ip ACL_SOFTPHONE trust-ipprec ip any any
cat6k-access> (enable) commit qos acl ACL_SOFTPHONE
cat6k-access> (enable) set qos acl map ACL_SOFTPHONE 7/1-48
```

Using Separate Access Layer Switches for IP Phones

You might want to connect the IP phones to separate switches in the wiring closet. This can avoid the need to upgrade your current data switches, and it serves to keep the voice and data networks completely separate. This type of installation is very similar to the scenario that uses separate ports on the wiring closet switch.

Speed and Duplex

Because there is no PC behind the IP phone in this type of installation, port speed and duplex settings are not as critical as in other types of installations. While it is safe to use the same configuration as with a single-cable connection (in case a PC is plugged into the second Ethernet port on the phone), this configuration is not required.

IP Addressing

The recommended configuration for connecting IP phones to separate access layer switches on a Cisco AVVID network is to use a separate IP address space and separate VLANs for IP telephony. In this case, the entire second switch, the newly installed VoIP-only Ethernet switch, runs a single VLAN. No trunking is necessary between the IP phone and the Ethernet switch. However, Cisco recommends that you use 802.1p for tagging VoIP packets from the IP phone as "important."

The commands for implementing this configuration are

```
cat4k> (enable) set port inlinepower 2/1-48 auto
cat4k> (enable) set port inlinepower 2/25-48 off
cat4k> (enable) set vlan 111 2/1-48
cat4k> (enable) set port host 2/1-48
cat4k> (enable) set port auxiliaryvlan 2/1-24 111
```

Classification and Queuing on the IP Phone

Because this type of installation puts the IP phones and any data PCs on separate physical cables, queuing on the IP phone is not required. However, since the IP phone is still a managed device, classification should still take place on the IP phone. This classification for VoIP packets can be handled in a variety of ways, depending upon which hardware is used in the wiring closet switch. If the wiring closet switch is a device that handles Layer 2 processing only, the CoS setting of the IP phone is used for classification at the access layer and into the distribution layer. The following example shows this type of configuration for a Catalyst 3500 XL or 2900 XL switch:

```
interface FastEthernet0/1
  switchport trunk encapsulation dot1q
  switchport trunk native vlan 112
  switchport mode trunk
  switchport voice vlan dot1p
  spanning-tree portfast
```

Summary

As described in this chapter, the following general guidelines and recommendations apply when connecting IP phones to the Cisco AVVID network:

- Use Auto-Negotiation for port settings on the wiring closet switch.
- Install IP phones on a separate voice-only subnet.
- Use PortFast to decrease IP phone boot time.
- Extend the classification trust boundary to the phone using **trust-ext** commands.
- Never allow PC applications to send traffic at a CoS or ToS value of 5-7.
- Use only Layer 3 or 4 wiring closet switches with SoftPhone.

In addition, the following caveats apply:

**Note**

Attaching IP phones to any shared media devices, such as an Ethernet hub, is not supported.

**Note**

Cascading (daisy chaining) of IP phones is not supported at this time.

Correctly connecting the IP phone is the first step in enabling QoS in the enterprise VoIP network. By enabling port Auto-Negotiation and classifying CoS and ToS, the IP phone can serve as the edge or boundary of the intelligent enterprise network.



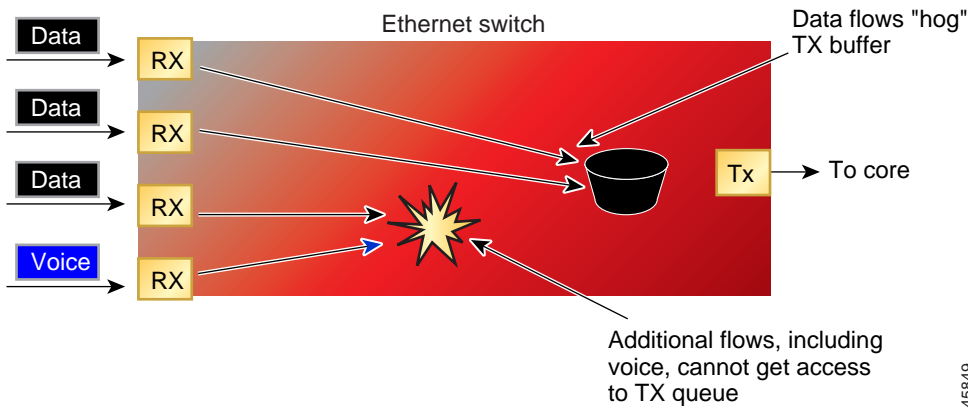
Designing a Campus

This chapter covers design considerations and recommendations for implementing the Cisco AVVID network in a campus environment.

Campus Switching Designs for Cisco AVVID

Until recently, the conventional wisdom was that Quality of Service (QoS) would never be an issue in the enterprise campus due to the bursty nature of network traffic and the capability of buffer overflow. Gradually, engineers have come to understand that buffering, not bandwidth, is the issue in the campus. For this reason, QoS tools are required to manage these buffers to minimize loss, delay, and delay variation. Figure 3-1 shows areas where transmit buffers can give rise to QoS issues.

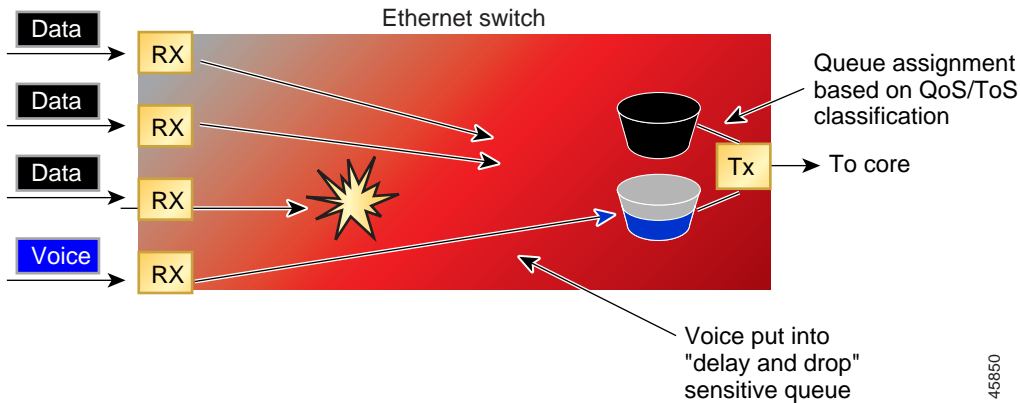
Figure 3-2 Loss of Voice Quality When Transmit Buffer Is Full



45849

VoIP traffic is sensitive to both delayed packets and dropped packets. As long as a campus is using Gigabit Ethernet trunks, which have extremely fast serialization times, delay should never be a factor regardless of the size of the queue buffer. Drops, however, always adversely affect voice quality in the campus. Using multiple queues on transmit interfaces is the only way to eliminate the potential for dropped traffic caused by buffers operating at 100% capacity. By separating voice and video (which are both sensitive to delays and drops) into their own queues, you can prevent flows from being dropped at the ingress interface even if data flows are filling up the data transmit buffer. Figure 3-3 illustrates the use of separate voice and data buffers.

Figure 3-3 Using Separate Transmit Buffers for Voice and Data



45850

**Note**

It is critical to verify that Flow Control is disabled when enabling QoS (multiple queues) on Catalyst switches. Flow Control will interfere with the configured queuing behavior by acting on the ports before queuing is activated. Flow Control is disabled by default.

Queue Scheduling

The scheduler process can use a variety of methods to service each of the transmit queues (voice and data). The easiest method is a Round-Robin (RR) algorithm, which services queue 1 through queue N in a sequential manner. While not robust, this is an extremely simple and efficient method that can be used for branch office and wiring closet switches. Distribution Layer switches use a Weighted Round-Robin (WRR) algorithm in which higher priority traffic is given a scheduling "weight."

Another option is to combine Round-Robin or Weighted Round-Robin scheduling with priority scheduling for applications that are sensitive to packet delay and drop. This uses a priority queue (PQ) that is always served first when there are packets in the queue. If there are no frames in the PQ, the additional queues are scheduled using RR or WRR.

Number of Queues

There has been much discussion about how many queues are actually needed on transmit interfaces in the campus. Should you add a queue to the wiring closet switches for each Class of Service (CoS) value? Should you add eight queues to the distribution layer switches? Should you add a queue for each of the 64 Differentiated Services Code Point (DSCP) values? This section presents some guidelines that address these questions.

First, it is important to remember that each port has a finite amount of buffer memory. A single queue has access to all the memory addresses in the buffer. As soon as a second queue is added, the finite buffer amount is split into two portions, one for each queue. Now all packets entering the switch must contend for a much smaller portion of buffer memory. During periods of high traffic, the buffer fills, and packets are dropped at the ingress interface. Because the majority of network traffic today is TCP-based, a dropped packet results in a re-send, which further increases network congestion. Therefore, queuing should be used cautiously and only when particular priority traffic is sensitive to packet delays and drops.

Two queues are adequate for wiring closet switches, where buffer management is less critical than at other layers. How these queues are serviced (Round-Robin, Weighted Round-Robin, or Priority Queuing) is less critical than the number of buffers because the scheduler process is extremely fast when compared to the aggregate amount of traffic.

Distribution layer switches require much more complex buffer management due to the flow aggregation occurring at that layer. Not only are priority queues needed, but you should also specify thresholds within the standard queues. Cisco has chosen to use multiple thresholds within queues instead of continually increasing the number of interface queues. As discussed earlier, each time a queue is configured and allocated, all of the memory buffers associated with that queue can be used only by frames meeting the queue entrance criteria. The following example illustrates this concept:

Assume that a Catalyst 4000 10/100 Ethernet port has two queues configured, one for VoIP (VoIP bearer and control traffic) and the default queue, which is used for Hypertext Transfer Protocol (HTTP), e-mail, File Transfer Protocol (FTP), logins, Windows NT Shares, and Network File System (NFS). The 128-KB voice queue is split into a 7:1 transmit and receive ratio. The transmit buffer memory is then further separated into high- and low-priority partitions in a 4:1 ratio. If the default traffic (the web, e-mail, and file shares) begins to congest the default queue, which is only 24 KB, packets begin dropping at the ingress interfaces. This occurs

regardless of whether the VoIP control traffic is using any of its queue buffers. The dropped packets of the TCP-oriented applications cause each of these applications to send the data again, aggravating the congested condition of the network. If this same scenario were configured with a single queue, but with multiple thresholds used for congestion avoidance, the default traffic would share the entire buffer space with the VoIP control traffic. Only during periods of congestion, when the entire buffer memory approaches saturation, would the lower priority traffic (HTTP and e-mail) be dropped.

This discussion does not imply that multiple queues are to be avoided in Cisco AVVID networks. As discussed earlier, the VoIP bearer streams *must* use a separate queue to eliminate the adverse affects that packet drops and delays have on voice quality. However, every single CoS or DSCP value should *not* get its own queue because the small size of the resulting default queue will cause many TCP re-sends and will actually increase network congestion.

In addition, the VoIP bearer channel is a bad candidate for queue congestion avoidance algorithms such as Weighted Random Early Detection (WRED). Queue thresholding uses the WRED algorithm to manage queue congestion when a preset threshold value is specified. Random Early Detection (RED) works by monitoring buffer congestion and discarding TCP packets if the congestion begins to increase. The result of the drop is that the sending endpoint detects the dropped traffic and slows the TCP sending rate by adjusting the window size. A WRED drop threshold is the percentage of buffer utilization at which traffic with a specified CoS value is dropped, leaving the buffer available for traffic with higher priority CoS values. The key is the word "Random" in the algorithm name. Even with weighting configured, WRED can still discard packets in any flow; it is just statistically more likely to drop them from the lower CoS thresholds.

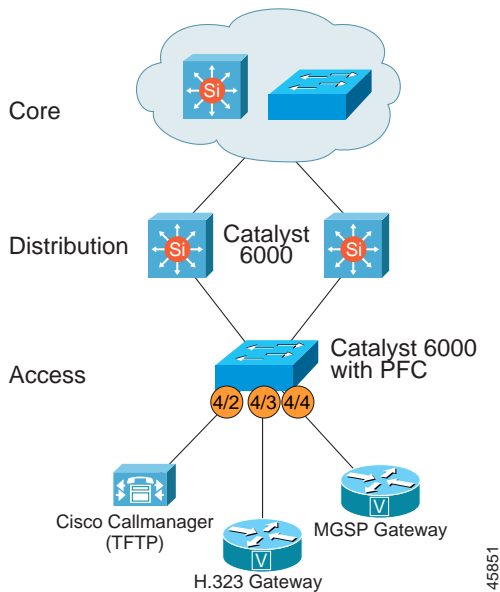
Marking Control and Management Traffic

In networks with high traffic loads, managing the delivery of control traffic is critical to ensuring a positive user experience with VoIP. An example where this comes into play is with the Delay to Dial-Tone (DTT) time. The Cisco IP Phones use Skinny Station Protocol to communicate with Cisco CallManager. When a Cisco IP Phone goes off hook, it "asks" Cisco CallManager what to do. Cisco CallManager then instructs the Cisco IP Phone to play dial-tone. If this Skinny Client Protocol management and control traffic is dropped or delayed within the network, the user experience is adversely affected. 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 the actual RTP stream), Access Control Lists (ACLs) are used to classify these streams on Catalyst 5000 and 6000 switches that are enabled for Layer 3/4. Examples of these configurations are detailed in the “Catalyst 6000 Access Layer” section on page 3-11. For designs where a Cisco IOS router is the first Layer 3 or 4 access point, ACLs are used. Examples of these configurations are included in Chapter 5, “Implementing a Wide Area Network.”

Figure 3-4 shows a typical server farm design, with an access layer switch providing access to the distribution layer.

Figure 3-4 Typical Server Farm



Skinny Protocol

Cisco CallManager communicates with IP phones and gateways using TCP ports 2000-2002. The following example commands classify all Skinny Protocol traffic from IP phones and gateways (VLAN 110) and Cisco CallManager (4/2) as DSCP 26 (AF31, which is backward compatible with IP Precedence 3).



Note

Beginning with Release 3.0(5), Cisco CallManager includes the ability to configure the CoS and ToS values for all VoIP control and management traffic from Cisco CallManager, the IP phones, and the Skinny Protocol gateways (this does not include the AT and AS model analog gateways). With this user-configurable classification, network element access lists are no longer required for marking Skinny Protocol VoIP control traffic. H.323 and Media Gateway Control Protocol (MGCP) traffic still require external, network element marking for several more months.

The following commands perform these functions:

1. Enable switch-wide QoS.
2. Create an access control list (ACL_IP-PHONES), marking all Skinny Client and Gateway Protocol traffic from the IP phones and from Skinny Protocol gateways with a DSCP value of 26 (AF31).
3. Add to the ACL_IP-PHONE access list, trusting all DSCP markings from the IP phone, so that the ToS=5 RTP traffic is not rewritten.
4. Create an access control list (ACL_VOIP_CONTROL), marking all Skinny Client and Gateway Protocol traffic from Cisco CallManager with a DSCP value of 26 (AF31).
5. Accept incoming Layer 2 CoS classification. (Current 10/100 version "1" line cards must have **trust-cos** enabled even though the parser returns an error).
6. Inform the port that all QoS associated with the port will be done on a VLAN basis.
7. Instruct the IP phone to rewrite CoS from the PC to CoS=0 within the IP phone Ethernet ASIC.
8. Inform Cisco CallManager port (4/2) that all QoS associated with the port will be done on a port basis.

9. Write the access control list to hardware.
10. Map the ACL_IP-PHONE access control list to the auxiliary VLAN.
11. Map the ACL_VOIP_CONTROL access control list to the Cisco CallManager port.

```
cat6k-access> (enable) set qos enable
cat6k-access> (enable) set qos acl ip ACL_IP-PHONES dscp 26 tcp any any range 2000 2002
cat6k-access> (enable) set qos acl ip ACL_IP-PHONES trust-cos ip any any
cat6k-access> (enable) set qos acl ip ACL_VOIP_CONTROL dscp 26 tcp any any range 2000 2002
cat6k-access> (enable) set port qos 5/1-48 trust trust-cos
cat6k-access> (enable) set port qos 5/1-48 vlan-based
cat6k-access> (enable) set port qos 5/1-48 trust-ext untrusted
cat6k-access> (enable) set port qos 4/2 port-based
cat6k-access> (enable) commit qos acl all
cat6k-access> (enable) set qos acl map ACL_IP-PHONES 110
cat6k-access> (enable) set qos acl map ACL_VOIP_CONTROL 4/2
```

H.323 Protocol

Cisco CallManager communicates with H.323 gateways using TCP ports 1720 (H.225) and 11xxx (H.245). The following example commands classify H.323 control traffic from Cisco CallManager (4/2) and from H.323 gateways (4/3) as DSCP 26 (AF31, which is backward compatible with IP Precedence 3).

```
cat6k-access> (enable) set qos acl ip ACL_VOIP_CONTROL dscp 26 tcp any any eq 1720
cat6k-access> (enable) set qos acl ip ACL_VOIP_CONTROL dscp 26 tcp any any range 11000
11999
cat6k-access> (enable) set port qos 4/2 port-based
cat6k-access> (enable) set port qos 4/3 port-based
cat6k-access> (enable) commit qos acl ACL_VOIP_CONTROL
cat6k-access> (enable) set qos acl map ACL_VOIP_CONTROL 4/2
cat6k-access> (enable) set qos acl map ACL_VOIP_CONTROL 4/3
```

MGCP

Cisco CallManager communicates with Media Gateway Control Protocol (MGCP) gateways using User Datagram Protocol (UDP) port 2427. The following example commands classify MGCP control traffic from Cisco CallManager (4/2) and from the MGCP gateway (4/4) as DSCP 26 (AF31, which is backward compatible with IP Precedence 3).

```
cat6k-access> (enable) set qos acl ip ACL_VOIP_CONTROL dscp 26 udp any any eq 2427
cat6k-access> (enable) set port qos 4/2 port-based
cat6k-access> (enable) set port qos 4/4 port-based
cat6k-access> (enable) commit qos acl ACL_VOIP_CONTROL
cat6k-access> (enable) set qos acl map ACL_VOIP_CONTROL 4/2
cat6k-access> (enable) set qos acl map ACL_VOIP_CONTROL 4/4
```

Example 3-1 shows the command and its associated output for verifying that the Access Control Lists (ACLs) are attached to the correct VLANs and ports.

Example 3-1 Verifying the ACLs

```
cat6k-access> (enable) sh qos acl map run all
```

ACL name	Type	Vlans
ACL_IP-PHONES	IP	110,111,112

ACL name	Type	Ports
ACL_IP-PHONES	IP	

ACL name	Type	Vlans
ACL_VOIP_CONTROL	IP	

ACL name	Type	Ports
ACL_VOIP_CONTROL	IP	4/2,4/3,4/4

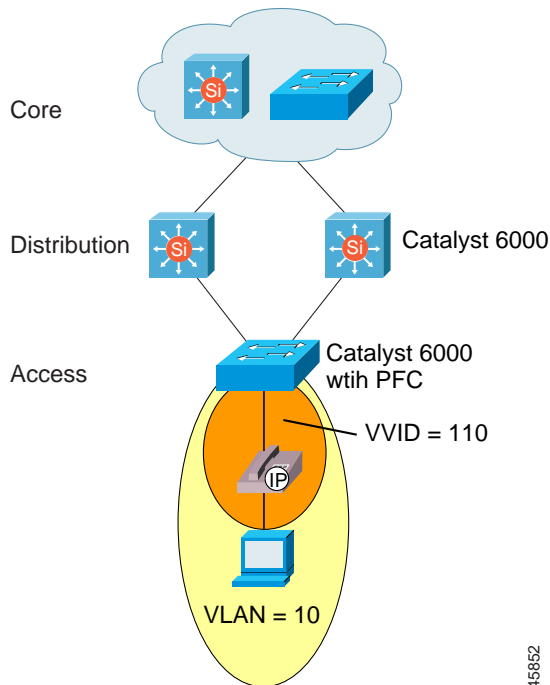
Catalyst 6000 Access Layer

One of the most popular campus configurations for Cisco AVVID solutions is to use Catalyst 6000 switches in both the wiring closet and the distribution and core layers. There are several compelling reasons for this:

- The Catalyst 6000 can provide in-line power to the IP phones.
- The Catalyst 6000 offers the highest growth potential.
- The Catalyst 6000 supports the most advanced Layer 2/3 campus QoS tools in the Cisco product line.

Figure 3-5 shows a general model for the Catalyst 6000 QoS configurations discussed in this guide.

Figure 3-5 General Model for Catalyst 6000 QoS Configurations



45862

With the addition of the Policy Feature Card (PFC) daughter card, the Catalyst 6000 is inherently capable of handling Layer 2, 3, and 4 QoS issues. The PFC can be used to enable advanced QoS tools such as packet classification and marking, scheduling, and congestion avoidance based on either Layer 2 or Layer 3 and 4 header information. Multiple receive and transmit queues with thresholds can be configured and used according to the QoS policy rules configured in the switch.

The Catalyst 6000 has two versions of Supervisor Engines, the Sup1 and Sup1A. There are also two versions of Catalyst 6000 line cards, the second of which is also denoted by an "A" product number. All Catalyst 6000 Ethernet modules support a single receive queue with four thresholds and two transmit queues, each with two thresholds. The "A" cards include enhanced QoS features that provide an additional priority queue for both ingress and egress interfaces. These queues are serviced in a Weighted Round-Robin (WRR) method, except for the priority queue, which is always serviced as soon as frames have entered the queue. To see how a port is configured, issue the **show port capabilities <mod/port>** CatOS command. The default QoS capabilities of the port can be changed using the **set qos map <port_type> rx | tx <queue#> <threshold#>** and **set qos wred-threshold** commands. When modifying the queue thresholds, it is important to remember that the higher priority queue has a higher numerical value.

Scheduling for the Catalyst 6000 transmit interfaces is managed by the WRR algorithm. Each queue is given a user-configurable weight. By default, the "high" queue is given 98% of the scheduler time, and the "low" queue is given just 2%. This ratio is conducive to ensuring that packets with a low delay tolerance are not delayed in a queue. This is also the reason behind giving the "low" queue a much higher percentage of the overall interface buffer. If the Priority Queue (PQ) is configured, it will always be serviced first. If no frames reside in the PQ, WRR begins to schedule the other two queues.

Catalyst 6000 Port Scheduling and Queuing Schemes

This section presents several suggested configurations for port scheduling and queuing on the Catalyst 6000.

Receive Interface

There are two possible configurations for the receive interface, depending on whether or not a priority queue is needed:

- 1Q4T

One standard queue with four drop thresholds.

8-KB receive buffer for 10/100 Mbps

64-KB receive buffer for 1000 Mbps

Available on all 10/100/1000 Mbps modules

The default values for the drop thresholds are

% of Buffer Capacity	Drop CoS Value
50%	0-1
60%	2-3
80%	4-5
100%	6-7

- 1P1Q4T

One Priority Queue (PQ) and one standard queue with four drop thresholds

Available only on certain versions of 10/100/1000 Mbps modules, depending on line card

By default, all CoS 5 frames are placed in the PQ, which uses a strict priority scheduling algorithm.

The default values for the drop thresholds in the standard queue are

Queue #	% of Buffer Capacity	Drop CoS Value
1	50%	0-1
1	60%	2-3
1	80%	4
1	100%	6-7
2	100%	5

Transmit Interface

There are two possible configurations for the transmit interface, depending on whether a priority queue is needed:

- 2Q2T

Two standard queues with two drop thresholds. The high-priority queue is allocated 20% of the total queue size, and the low-priority queue is allocated 80% of the total queue size.

Available on all 10/100/1000 Mbps modules

The default values for the drop thresholds are

Queue #	% of Buffer Capacity	Drop CoS Value
1 - Low Priority - 80% of total queue size	40%	0-1
	100%	2-3
2 - High Priority - 20% of total queue size	40%	4-5
	100%	6-7

- 1P2Q2T

One Priority Queue (PQ) and two standard queues with two drop thresholds. By default, all CoS 5 frames are placed in the PQ, which uses a strict priority scheduling algorithm that always services the PQ first, and, once the PQ is

empty, WRR is used on the remaining queues. The PQ gets allocated 15% of the total queue size, as does the high-priority queue. The low-priority queue is allocated 70% of the total queue size.

Available only on certain versions of 10/100/1000 Mbps modules, depending on line card

The default values for the drop thresholds are

Queue #	% of Buffer Capacity	Drop CoS Value
1 - Low Priority - 70% of total queue size	40%	0-1
	100%	2-3
2 - High Priority - 15% of total queue size	40%	4
	100%	6-7
3 - Priority Queue - 15% of total queue size	100%	5

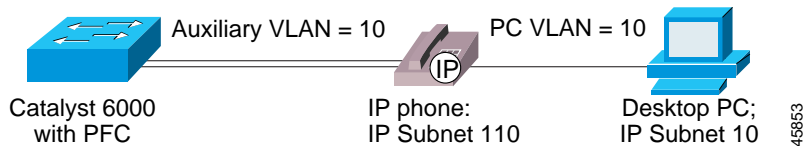
Configuring QoS Parameters

After you have connected the IP phone to the wiring closet switch (see Chapter 2, “Connecting IP Phones”), it is time to configure the QoS parameters on the switch. This includes setting up multiple queues on all ports, configuring access to the queues, setting thresholds for traffic drops, and connecting the switch to the distribution or core layer. The following sections detail these steps.

IP Phone Port Queuing

If you use a single cable to connect IP phone, as illustrated in Figure 3-6, the access port is configured to trust the IP phone and not the attached PC. The port is also configured to use multiple transmit queues, one being a priority queue for voice traffic.

Figure 3-6 Using a Single Cable to Connect an IP Phone



The following commands enable QoS on the access layer Catalyst 6000 by performing these functions:

1. Enable switch-wide QoS.
2. Inform the port that all QoS associated with the port will be done on a VLAN basis.
3. Instruct the IP phone to rewrite CoS from the PC to CoS=0 within the IP phone Ethernet ASIC.
4. Accept incoming Layer 2 CoS classification. (Current 10/100 version "1" line cards must still have **trust-cos** enabled even though the parser returns an error).
5. Create an access list that accepts incoming Layer 3 ToS classification (necessary only on 10/100 ports).
6. Write the access list to hardware.
7. Map the access list to the auxiliary VLAN.

```

cat6k-access> (enable) set qos enable
cat6k-access> (enable) set port qos 5/1-48 vlan-based
cat6k-access> (enable) set port qos 5/1-48 trust-ext untrusted
cat6k-access> (enable) set port qos 5/1-48 trust trust-cos
cat6k-access> (enable) set qos acl ip ACL_IP-PHONES trust-cos any
cat6k-access> (enable) commit qos acl ACL_IP-PHONES
cat6k-access> (enable) set qos acl map ACL_IP-PHONES 110

```

Once QoS has been enabled on the Catalyst 6000 access layer switch, you can use the following command to place all CoS=3 (VoIP control) traffic into the second transmit queue, with a low drop threshold, to ensure successful call control during periods of heavy congestion. All CoS=5 (VoIP RTP Bearer) traffic is placed into the second queue automatically.

```
cat6k-access> (enable) set qos map 2q2t tx 2 1 cos 3
```

Verifying IP Phone Access Port Configuration

One of the fundamental processes of implementing Quality of Service (QoS) is verifying that the configurations are actually performing as expected. On the Catalyst 6000 access layer switch, you can verify configuration performance during periods of high congestion by examining the output of the following commands:

- `show port qos <mod/port>`
This command shows the QoS settings for the specified port. See Example 3-2.
- `show qos info runtime <mod/port>`
This command shows QoS runtime information for the specified port. See Example 3-3.
- `show mac <mod/port>`
This command shows Media Access Control (MAC) information for the specified port. See Example 3-4.
- `show qos statistics l3`
This command shows summary QoS statistics for all ports. See Example 3-5.
- `show qos stat <mod/port>`
This command shows detailed QoS statistics for the specified port. See Example 3-6.

Example 3-2 Displaying QoS Settings

```
cat6k-access> (enable) sh port qos 5/1
QoS is enabled for the switch
QoS policy source for the switch set to local.
```

Port	Interface	Type	Interface	Type	Policy	Source	Policy	Source
	config		runtime		config		runtime	
5/1	vlan-based		vlan-based			COPS		local

Port	TxPort	Type	RxPort	Type	Trust	Type	Trust	Type	Def	CoS	Def	CoS
					config		runtime		config		runtime	
5/1		2q2t		1q4t	trust-cos		trust-cos*			0		0

Port	Ext-Trust	Ext-Cos
5/1	untrusted	0

(*)Runtime trust type set to untrusted.

Config:

Port	ACL name	Type
No ACL is mapped to port 5/1.		
ACL is mapped to VLAN		

Runtime:

Port	ACL name	Type
No ACL is mapped to port 5/1.		

Example 3-3 Displaying QoS Runtime Information

```

cat6k-access>(enable) sh qos info run 5/1
Run time setting of QoS:
QoS is enabled
Policy Source of port 5/1: Local
Current 10/100 "1" linecards support 2q2t/1q4t only
Tx port type of port 5/1 : 2q2t
Rx port type of port 5/1 : 1q4t
Interface type: vlan-based
ACL is mapped to VLAN
ACL attached:
The qos trust type is set to trust-cos.
Warning: Runtime trust type set to untrusted.
Default CoS = 0
Queue and Threshold Mapping for 2q2t (tx):
Queue Threshold CoS
-----
1      1      0 1
1      2      2
2      1      3 4 5
2      2      6 7
Queue and Threshold Mapping for 1q4t (rx):
Queue Threshold CoS
-----
1      1      0 1
1      2      2
1      3      3 4 5
1      4      6 7
. . .

```

Example 3-4 Displaying MAC Information

```
cat6k-access> (enable) sh mac 5/1
```

```
Port      Rcv-Unicast      Rcv-Multicast      Rcv-Broadcast
-----
5/1              267223              37                  4

Port      Xmit-Unicast      Xmit-Multicast      Xmit-Broadcast
-----
5/1              28748894            5206                72

Port      Rcv-Octet      Xmit-Octet
-----
5/1              17178128            1840430081

"Out-Discards" are packets drooped due to congestion in the tx interface buffers
MAC      Dely-Exced  MTU-Exced  In-Discard  Out-Discard
-----
5/1      0           0           0           262140
```

Example 3-5 Displaying QoS Summary Statistics

```
cat6k-access> (enable) sh qos stat l3
VoIP Control packets that have been re-written with CoS=3/DSCP=26 (AF31)
Packets dropped due to policing:          0
IP packets with ToS changed:             1885
IP packets with CoS changed:             781
Non-IP packets with CoS changed:         0
```

Example 3-6 Displaying QoS Detailed Statistics

```
cat6k-access> (enable) sh qos stat 5/1
All packets dropped are in the 1st drop threshold of queue #1
Tx port type of port 5/1 : 2q2t
Q #  Threshold #:Packets dropped
---
1    1:393210 pkts, 2:0 pkts
2    1:0 pkts, 2:0 pkts
Rx port type of port 5/1 : 1q4t
Q #  Threshold #:Packets dropped
---
1    1:0 pkts, 2:0 pkts, 3:0 pkts, 4:0 pkts
```

Uplink Interface to the Distribution Switch

Once you have configured all the access port queuing, you must also configure the uplink interfaces to the distribution/core switch. This involves enabling trust for Ethernet frames coming into the trunk port (1/1 in this example), manipulating the CoS-to-queue mapping entrance criteria, and mapping the CoS and IP Precedence values to the appropriate DSCP value. The procedure for doing this is outlined in the following sections.

MLS and Catalyst QoS Configuration

If the IP phones are in a different VLAN than Cisco CallManager, additional configuration is required. Any time a packet is sent to the Multilayer Switch Feature Card (MSFC) for Layer 3 switching, the CoS is set to 0. Because most configurations have the MSFC located in the distribution layer switch, the access layer switch must trust all DSCP tagging on the uplink trunk from the distribution layer. This enables the DSCP marking to be retained and used for DSCP-to-CoS Layer 3 classification in the wiring closet switch. Use **trust-cos** for Layer 2 uplinks and **trust-dscp** for Layer 3 uplinks; for example:

```
cat6k-access> (enable) set port qos 1/1 trust trust-dscp
```

Catalyst 6000 Transmit Queue Configuration

All VoIP (CoS=5) traffic will be placed into the egress interface Priority Queue on 1p2q2t interfaces and Queue 2 on 2q2t interfaces as soon as you enable QoS. However, you must perform the additional step of configuring the Catalyst 6000 CoS queue admission rules to ensure that CoS=3 (VoIP control) traffic is placed into the second queue. Use the following commands to perform this configuration:

```
cat6k-access> (enable) set qos map 1p2q2t tx 2 1 cos 3
cat6k-access> (enable) set qos map 2q2t tx 2 1 cos 3
```

Catalyst 6000 CoS/ToS-to-DSCP Mapping Configuration

Cisco follows the Internet Engineering Task Force (IETF) recommendations for setting the DSCP classification values for both the VoIP control plane traffic and VoIP bearer or media plane traffic. The recommended settings are DSCP=AF31 for VoIP control plane and DSCP=EF for VoIP bearer plane. To map the Layer 2 CoS and Layer 3 IP precedence settings correctly to these DSCP values, you must modify the default CoS/ToS-to-DSCP mappings as follows:

```
cat6k-distrib> (enable) set qos cos-dscp-map 0 8 16 26 32 46 48 56
cat6k-distrib> (enable) set qos ipprec-dscp-map 0 8 16 26 32 46 48 56
```

Verifying CoS/ToS-to-DSCP Mapping

To verify that the CoS and ToS settings are mapped correctly to the DSCP values, use the following two commands (shown with their associated output):

```
cat6k-distrib> (enable) sh qos map run cos-dscp-map
CoS - DSCP map:
CoS    DSCP
---    ----
  0     0
  1     8
  2    16
  3    26 -> 26 = AF31
  4    32
  5    46 -> 46 = EF
  6    48
  7    56

cat6k-distrib> (enable) sh qos map run ipprec-dscp-map
IP-Precedence - DSCP map:
IP-Prec  DSCP
-----  ----
  0     0
  1     8
  2    16
  3    26 -> 26 = AF31
  4    32
  5    46 -> 46 = EF
  6    48
  7    56
```

Catalyst 4000 Access Layer

Another popular campus configuration for Cisco AVVID networks uses Catalyst 2948G, 2980G, and 4000 series switches in the wiring closets. There are several compelling reasons for this:

- The Catalyst 4006 can provide in-line power to the IP phones.
- The Catalyst 4000 offers a very low price per port.
- These switches provide extremely scalable, high-speed switching.

Starting with CatOS Release 5.2, the Catalyst 4000 lines support dual-transmit queues on every interface. Admission to the queues is based on Layer 2 CoS markings and is configurable in 802.1p User Priority pairs.

Catalyst 4000 Port Scheduling and Queuing Schemes

This section presents several suggested configurations for port scheduling and queuing on the Catalyst 4000.

Receive Interface

The recommended configuration for the receive interface is

- FIFO
One standard FIFO (First-In, First-Out) queue.

Transmit Interface

The recommended configuration for the transmit interface is

- 2QIT

Two standard queues with a single threshold. Scheduling is done on a Round-Robin (RR) basis. Admission to the queues is based on 802.1p CoS value and is user configurable in pairs. If you enable QoS but do not modify the CoS-to-transmit queue mappings, switch performance could be affected because all traffic is assigned to queue 1.



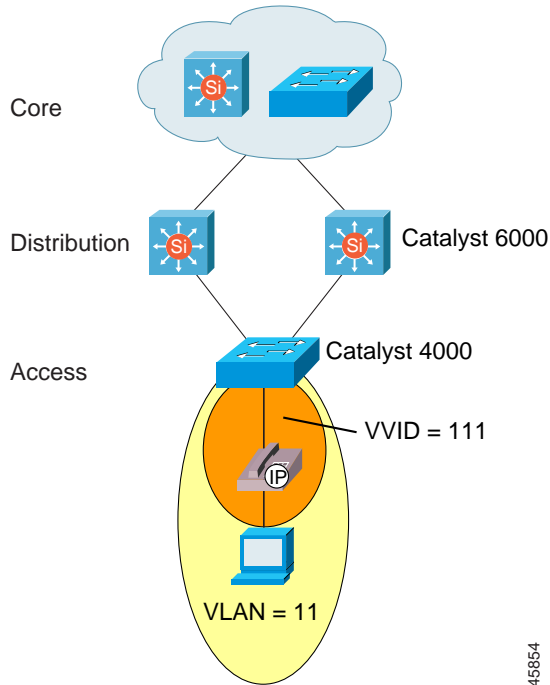
Note Once QoS is enabled on the Catalyst 4000, you must change the CoS mappings to utilize the newly created queue.

The default queue admission criteria for the Catalyst 4000 are

Queue #	Queue Admission CoS Value
1	0-7
2	Broadcast, Multicast, and Unknown Traffic

Figure 3-7 shows a general model for the Catalyst 4000 QoS configurations discussed in this guide.

Figure 3-7 General Model for Catalyst 4000 QoS Configurations



Catalyst 4000 Switch-Wide QoS

By default, only one queue is enabled on the Catalyst 4000 line of switches. Use the **set qos map** commands to enable the use of the second queue in CatOS Release 5.5.1. VoIP Control (CoS=3) frames should be placed into the second queue in the Catalyst 4000. These maps must be configured in pairs of CoS values because the Catalyst 4000 examines only the first two CoS bits; for example:

```
cat4k> (enable) set qos enable
cat4k> (enable) set qos map 2q1t 1 1 cos 0-1
cat4k> (enable) set qos map 2q1t 2 1 cos 2-3
cat4k> (enable) set qos map 2q1t 2 1 cos 4-5
cat4k> (enable) set qos map 2q1t 2 1 cos 6-7
```

Verifying Catalyst 4000 Queue Admission Configuration

To verify the queue admission configuration on the Catalyst 4000, use the following command (shown with its associated output):

```
cat4k> (enable) show qos info runtime
Run time setting of QoS:
QoS is enabled
All ports have 2 transmit queues with 1 drop thresholds (2q1t).
Default CoS = 0
Queue and Threshold Mapping:
Queue Threshold CoS
-----
1      1      0 1
2      1      2 3 4 5 6 7
```

IP Phone Port Queuing

In CatOS Release 5.5.1, the Catalyst 4000 line does not offer any advanced IP phone queuing features. Because of this, the Catalyst 4000 depends on the default CoS marking and enforcement on the IP phone. For more details, see Chapter 2, “Connecting IP Phones.”

Uplink Interface to the Distribution Switch

No special queuing or scheduling commands need to be configured on the Catalyst 4000 side of the link (from the access layer Catalyst 4000 to the distribution layer Catalyst 6000) because queuing is automatically enabled once QoS has been enabled and classification and queue admission have been configured.

You can perform additional uplink configuration if you are using the Catalyst 4000 with the Layer 3 engine (the WS-X4232, which enables IP, IPX, and Multicast routing for the switch). The Layer 3 engine enables the Catalyst 4000 to support four transmit queues based on IP precedence for

entrance criteria on the two-gigabit uplinks. The four queues are scheduled using a user-configurable WRR algorithm. In this case, the transmit interface configuration is as follows:

- 4QIT

Four standard queues with a single threshold. Scheduling is done on a Round-Robin (RR) basis. Admission to the queues is based on 802.1p CoS value and is user configurable in pairs.



Note Once QoS is enabled on the Catalyst, you must change CoS mappings to utilize the newly created queue. Note that the Layer 3 queue numbering is the reverse of the Layer 2 numbering.

The default Layer 3 1000-Mbps uplink queue admission criteria for the Catalyst 4000 are as follows:

Queue #	Queue Admission IP Precedence Value
1	6-7
2	4-5
3	2-3
4	0-1

Catalyst 3500 Access Layer

The Cisco AVVID features in the Catalyst 2900 and Catalyst 3500 series, with a minimum Cisco IOS release of 12.0(5)XU, allow interaction with the IP phones for extending the CoS marking rules. In addition, the Catalyst 2900 XL and 3500 XL switches can classify untagged packets at the ingress ports by setting a default CoS priority for each port. However, these switches (except for the 3548 XL) cannot reclassify any tagged packets, and they honor only the 802.1p priority and place the packets in the appropriate transmit queue. All Catalyst 3500 switches and all Catalyst 2900 XL switches with 8-MB DRAM support these QoS features. The Catalyst 2900 XL with 4-MB DRAM does not support QoS features.

Catalyst 3500 Port Scheduling and Queuing Schemes

This section presents several suggested configurations for port scheduling and queuing on the Catalyst 3500.

Receive Interface

The recommended configuration for the receive interface is

- 1Q-FIFO
One standard FIFO (First-In, First-Out) queue.

Transmit Interface 10/100 Ports

The recommended configuration for the transmit interface for 10/100 ports is

- 2QIT
Two standard queues with a single drop threshold. Scheduling is done on a priority-scheduling basis. Admission to the queues is based on 802.1p CoS or port-priority CoS value and is *not* user configurable.

The queue admission criteria for the Catalyst 3500 are as follows:

Queue #	Queue Admission CoS Value
1	0-3
2	4-7

Transmit Interface Gigabit Ethernet Ports

The recommended configuration for the transmit interface for gigabit Ethernet ports is

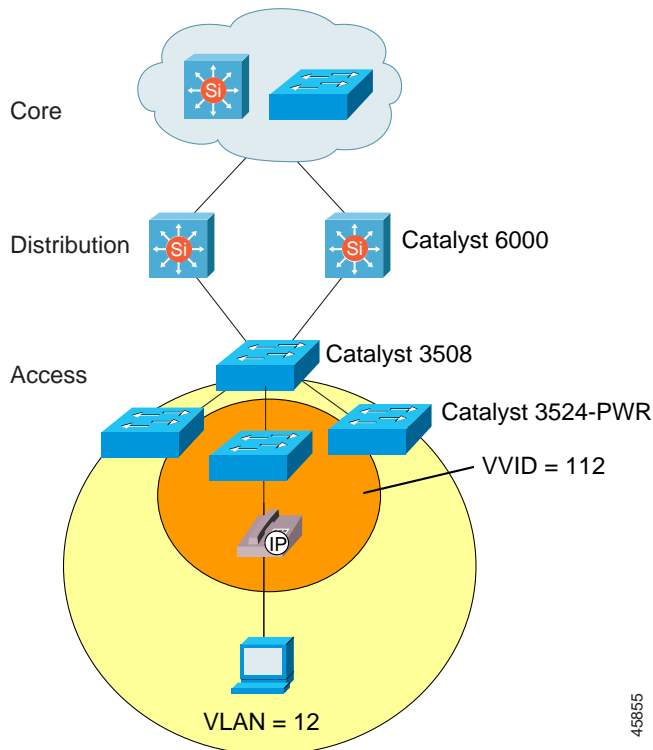
- 8Q-FIFO
Eight standard queues with a single drop threshold. Currently, only two queues are used. Scheduling is done on a priority-scheduling basis. Admission to the queues is based on 802.1p or port-priority CoS values and is *not* user configurable.

The gigabit Ethernet queue admission criteria are as follows:

Queue #	Queue Admission CoS Value
1	0-3
2	4-7
3-8	Not Used

Figure 3-8 shows a general model for the Catalyst 3500 QoS configurations discussed in this guide.

Figure 3-8 General Model for Catalyst 3500 QoS Configurations



IP Phone Port Queuing

If you use a single cable to install an IP phone, the access port is configured to trust the IP phone and not the attached PC. The port is also configured to use multiple transmit queues on all interfaces.

The commands to configure IP phone port queuing are

```
interface FastEthernet0/1
  switchport trunk encapsulation dot1q
  switchport trunk native vlan 12
  switchport mode trunk
  switchport voice vlan 112
  switchport priority extend cos 0
  spanning-tree portfast
```

Uplink Interface to the Distribution Switch

The recommended design for wiring closet configurations of Catalyst 3500 XL series switches is a star topology with a Catalyst 3524 PWR XL connected to a Catalyst 3508, which has dual uplinks to the distribution layer Catalyst 6000 switches. These uplinks are gigabit Ethernet links load balancing VLANs across the uplinks and configured with UplinkFast for fast Layer 2 convergence.

**Note**

A Catalyst 3500 series GigaStack configuration cannot provide guaranteed voice QoS because it is essentially a shared media access model.

The commands for this configuration are

```
interface GigabitEthernet0/1
  switchport trunk encapsulation dot1q
  switchport mode trunk
```

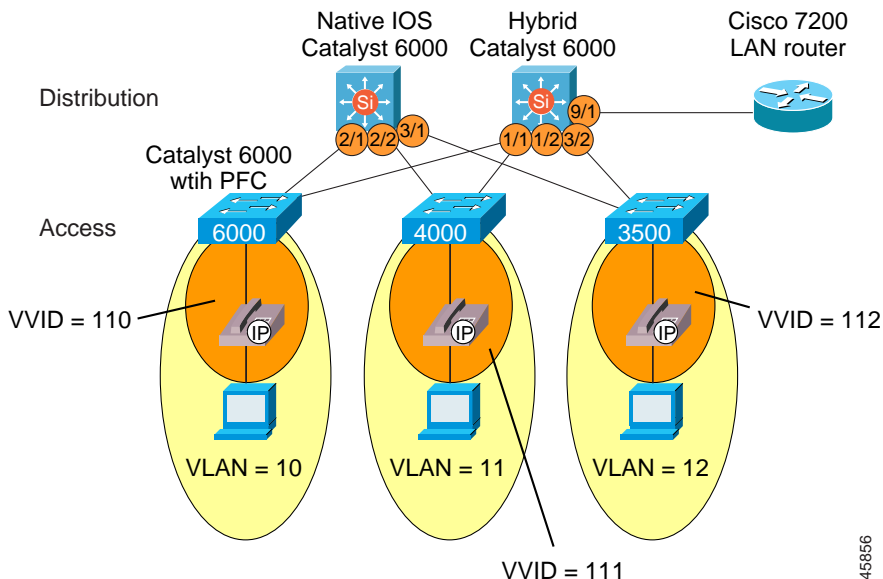
Catalyst 6000 Distribution Layer

After you configure the access switch and attach it to the distribution layer, you must set up Quality of Service (QoS) on the distribution switches. This requires the following changes to the configuration of the distribution switches:

- Configure VoIP control traffic transmit queuing.
- Configure the distribution layer with a Layer 3 access switch:
 - Enable trust ToS and DSCP from the access layer.
 - Configure ToS-to-DSCP mappings.
- Configure the distribution layer with a Layer 2 access switch:
 - Enable trust CoS and DSCP from the access layer.
 - Configure CoS-to-DSCP mappings.
 - Configure Layer 3 access lists for VoIP control traffic classification.
- Configure the connection to the Cisco 7200 WAN router.

Figure 3-9 shows a general model for these Catalyst 6000 distribution layer configurations, which are discussed in the following sections.

Figure 3-9 General Model for Catalyst 6000 Distribution Layer Configurations



45856

Configuring Catalyst 6000 Distribution Layer VoIP Control Traffic Transmit Queue

As soon as QoS is enabled, all VoIP (CoS=5) traffic is placed into the egress interface Priority Queue on 1p2q2t interfaces and into Queue #2 on 2q2t interfaces (for all versions "1" of the 10/100 line cards). You must also perform an additional step of configuring the Catalyst 6000 CoS queue admission rules to ensure that CoS=3 traffic flows (VoIP control traffic) are placed into the second queue.

The commands for performing this configuration are

```
cat6k-distrib> (enable) set qos map 1p2q2t tx queue 2 1 cos 3
cat6k-distrib> (enable) set qos map 2q2t tx queue 2 1 cos 3
```

Catalyst 6000 Distribution Layer Configuration with a Catalyst 6000-PFC Access Layer

Once you have enabled QoS on the distribution layer switch and have modified the default queue admission, two configuration steps remain for completing the integration with an access layer switch that is enabled for Layer 3:

- Enable trust DSCP from the access layer.
- Configure ToS-to-DSCP mappings.

Trust DSCP from the Layer 3 Access Switch

Enable trust for DSCP values from adjacent Layer 3 access switches. Use port-base QoS on the trunking port and use **trust-dscp** instead of **trust-cos**. This is because `trust-cos` overwrites the Layer 3 DSCP value with the mapped CoS, and there is no need to do this since classification is done at the access layer.

The commands for this configuration are

```
cat6k-distrib> (enable) set port qos 1/1 port-based
cat6k-distrib> (enable) set port qos 1/1 trust trust-dscp
```

Catalyst 6000 ToS-to-DSCP Mapping Configuration

Cisco follows the Internet Engineering Task Force (IETF) recommendations for setting the DSCP classification values for both the VoIP control plane traffic and VoIP bearer or media plane traffic. The recommended settings are DSCP=AF31 for VoIP control plane and DSCP=EF for VoIP bearer plane. To map the Layer 3 IP precedence settings correctly to these DSCP values, you must modify the default ToS-to-DSCP mappings as follows:

```
cat6k-distrib> (enable) set qos ipprec-dscp-map 0 8 16 26 32 46 48 56

cat6k-distrib> (enable) sh qos map run ipprec-dscp-map
IP-Precedence - DSCP map:
IP-Prec      DSCP
-----      -
          0      0
          1      8
          2     16
          3     26 -> 26 = AF31
          4     32
          5     46 -> 46 = EF
          6     48
          7     56
```

Catalyst 6000 Distribution Layer Configuration with an Access Switch Enabled for Layer 2 Only

Once you have enabled QoS on the distribution layer switch and have modified the default queue admission, you must perform three additional configuration steps to complete the integration with a Layer 2 access switch:

- Enable trust CoS from the access layer.
- Configure CoS-to-DSCP mappings.
- Configure Layer 3 access lists for VoIP control traffic classification. (See the “Marking Control and Management Traffic” section on page 3-6.)

Trust CoS from the Layer 2 Access Switch

Enable trust for CoS values from adjacent Layer 2 access switches. Use **vlan-based** QoS on the trunking port and use **trust-cos** instead of **trust-dscp**. This configuration is used when the access layer switch is only a Layer 2 device doing CoS classification.

The command for this configuration is

```
cat6k-distrib> (enable) set port qos 1/2,3/2 vlan-based
cat6k-distrib> (enable) set port qos 1/2,3/2 trust trust-cos
```

Catalyst 6000 CoS-to-DSCP Mapping Configuration

Cisco follows the IETF recommendations for setting the DSCP classification values for both the VoIP control plane traffic and VoIP bearer or media plane traffic. The recommended settings are DSCP=AF31 for VoIP control plane and DSCP=EF for VoIP bearer plane. To map the Layer 2 settings correctly to these DSCP values, you must modify the default CoS-to-DSCP mappings as follows:

```
cat6k-distrib> (enable) set qos cos-dscp-map 0 8 16 26 32 46 48 56
```

```
cat6k-distrib> (enable) sh qos map run cos-dscp-map
```

```
CoS - DSCP map:
CoS   DSCP
---   ----
  0    0
  1    8
  2   16
  3   26 -> 26 = AF31
  4   32
  5   46 -> 46 = EF
  6   48
```

Configuring Layer 3 Access Lists for VoIP Control Traffic Classification

To configure Layer 3 access lists for VoIP Control traffic classification, use the following commands (shown with their associated output). Also use the `ACL_IP-PHONES` access list from the “Marking Control and Management Traffic” section on page 3-6.

```
cat6k-distrib> (enable) set port qos 1/2,3/2 vlan-based
cat6k-distrib> (enable) set qos acl map ACL_IP-PHONES 111
```

```
cat6k-distrib> (enable) sh qos acl map run ACL_IP-PHONES
ACL name                               Type Vlans
-----
ACL_IP-PHONES                           IP 110,111,112
ACL name                               Type Ports
-----
ACL_IP-PHONES                           IP
```

```
cat6k-distrib> (enable) sh qos acl info run ACL_IP-PHONES
```

```
set qos acl IP ACL_IP-PHONES
-----
1. dscp 26 tcp any any range 2000 2002
2. dscp 26 tcp any any eq 1720
3. dscp 26 tcp any any range 11000 11999
4. dscp 26 udp any any eq 2427
5. trust-cos any
```



Note

Beginning with Release 3.0(5), Cisco CallManager includes the ability to configure the CoS and ToS values for all VoIP control and management traffic from Cisco CallManager, the IP phones, and the Skinny Protocol gateways (this does not include the AT and AS model analog gateways). With this user-configurable classification, network element access lists are no longer required for marking Skinny Protocol VoIP control traffic. H.323 and Media Gateway Control Protocol (MGCP) traffic still require external, network element marking for several more months.

Configuring the Connection to the Cisco 7200 WAN Router

Use the following commands to configure the connection to the Cisco 7200 WAN router.



Note

Current 10/100 version "1" line cards must still have **trust-ipprec** enabled even though the parser returns an error

```
cat6k-distrib> (enable) set port qos 9/1 port-based
cat6k-distrib> (enable) set port qos 9/1 trust trust-ipprec
```

```
cat6k-distrib> (enable) set qos acl ip ACL_TRUST-WAN trust-ipprec any
cat6k-distrib> (enable) commit qos acl ACL_TRUST-WAN
cat6k-distrib> (enable) set qos acl map ACL_TRUST-WAN 9/1
```

```
cat6k-distrib> (enable) sh port qos 9/1
QoS is enabled for the switch.
QoS policy source for the switch set to local.
```

Port	Interface	Type	Interface	Type	Policy	Source	Policy	Source
	config		runtime		config		runtime	
9/1	port-based		port-based			COPS		local

Port	TxPort	Type	RxPort	Type	Trust	Type	Trust	Type	Def	CoS	Def	CoS
					config		runtime		config	runtime	config	runtime
9/1		2q2t	1q4t		trust-ipprec		trust-ipprec		0		0	

Port	Ext-Trust	Ext-Cos
9/1	untrusted	0

(*)Runtime trust type set to untrusted.

Config:

Port	ACL name	Type
9/1	ACL_TRUST-WAN	IP

Runtime:

Port	ACL name	Type
9/1	ACL_TRUST-WAN	IP

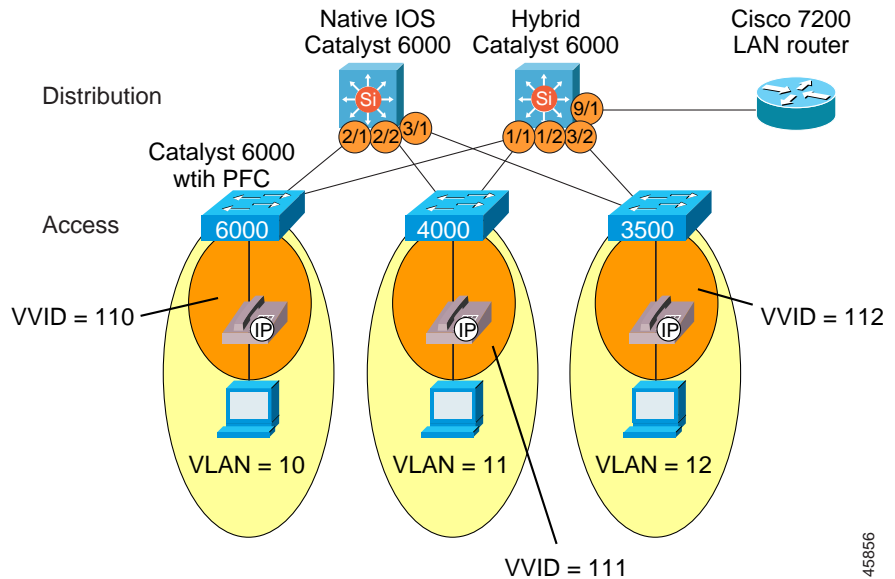
Catalyst 6000 Distribution/Core Running Native IOS

After you configure the access switch and attach it to the distribution layer, you must set up Quality of Service (QoS) on the distribution switches. This requires the following changes to the configuration of the distribution switches:

- Configure QoS.
- Configure VoIP control traffic transmit queuing.
- Configure the distribution layer with a Layer 3 access switch:
 - Enable trust ToS and DSCP from the access layer.
 - Configure ToS-to-DSCP mappings.
- Configure the distribution layer with a Layer 2 access switch:
 - Enable trust CoS and DSCP from the access layer.
 - Configure CoS-to-DSCP mappings.
 - Configure the QoS policies and Layer 3 access lists for VoIP control traffic classification.

Figure 3-10 shows a general model for configurations running Native Cisco IOS on Catalyst 6000 distribution layer switches. The configuration details are discussed in the sections that follow.

Figure 3-10 General Model for Native Cisco IOS Running on Catalyst 6000 Distribution Layer Switches



45856

Configuring QoS on the Native Cisco IOS Catalyst 6000

To enable QoS on the Catalyst 6000 with Native Cisco IOS, use the following command:

```
m1s qos
```

Configuring Transmit Queue Admission for VoIP Control Traffic

As soon as QoS is enabled, all VoIP (CoS=5) traffic is placed into the egress interface Priority Queue on 1p2q2t interfaces and into Queue #2 on 2q2t interfaces (for all versions "1" of the 10/100 line cards). You must also perform an additional step of configuring the Catalyst 6000 CoS queue admission rules to ensure that CoS=3 traffic flows (VoIP control traffic) are placed into the second queue.

The commands for performing this configuration are

```
int range gigabitEthernet 1/1 - 2
  wrr-queue cos-map 1 2 2
  wrr-queue cos-map 2 1 3 4
```

Catalyst 6000 Native Cisco IOS Distribution Layer Configuration with a Catalyst 6000-PFC Access Layer

Once you have enabled QoS on the Native Cisco IOS distribution layer switch and have modified the default queue admission, two configuration steps remain for completing the integration with an access layer switch that is enabled for Layer 3:

- Enable trust DSCP from the access layer.
- Configure ToS-to-DSCP mappings.

Trust DSCP from the Layer 3 Access Switch

Enable trust for DSCP values from adjacent Layer 3 access switches. Use port-base QoS on the trunking port (port-based QoS is enabled by default when **mls qos** is configured), and use **mls qos trust dscp** instead of the CatOS **trust-dscp**.



Note

Classification has already been established at the access layer in this model.

The commands for this configuration are

```
interface GigabitEthernet2/1
  description trunk port to PFC enabled cat6k-access
  no ip address
  wrr-queue cos-map 1 2 2
  wrr-queue cos-map 2 1 3 4
  mls qos trust dscp
  switchport
  switchport trunk encapsulation dot1q
  switchport mode trunk
```

Native Cisco IOS ToS-to-DSCP Mapping Configuration for Layer 3 Access Switches

Cisco follows the IETF recommendations for setting the DSCP classification values for both the VoIP control plane traffic and VoIP bearer or media plane traffic. The recommended settings are DSCP=AF31 for VoIP control plane and DSCP=EF for VoIP bearer plane. To map the Layer 3 IP precedence settings correctly to these DSCP values, you must modify the default ToS-to-DSCP mappings.



Note

The Catalyst 6000 numerical values of 26 and 46 correlate to DSCP=AF31 and DSCP=EF, respectively. This is done in global configuration mode.

The command for this configuration is

```
mls qos map ip-prec-dscp 0 8 16 26 32 46 56 0
```

Catalyst 6000 Native Cisco IOS Distribution Layer Configuration with an Access Switch Enabled for Layer 2 Only

Once you have enabled QoS on the distribution layer switch and have modified the default queue admission, you must perform three additional configuration steps to complete the integration with a Layer 2 access switch:

- Enable trust CoS from the access layer.
- Configure CoS-to-DSCP mappings.
- Configure Layer 3 access lists for VoIP control traffic classification. (See the “Marking Control and Management Traffic” section on page 3-6.)

Trust CoS from the Layer 2 Access Switch

Enable trust for CoS values from adjacent Layer 2 access switches. Use port-base QoS on the trunking port, and use the Native IOS command **mls qos trust cos** instead of CatOS **trust-cos**. This configuration is used when the access layer switch is only a Layer 2 device doing CoS classification.

The commands for this configuration are

```
interface GigabitEthernet2/2
  description trunk port to layer 2-only cat4k
  no ip address
  wrp-queue cos-map 1 2 2
  wrp-queue cos-map 2 1 3 4
  mls qos vlan-based
  mls qos trust cos
  switchport
  switchport trunk encapsulation dot1q
  switchport mode trunk
!
interface GigabitEthernet3/1
  description trunk port to layer 2-only 3500
  no ip address
  wrp-queue cos-map 1 2 2
  wrp-queue cos-map 2 1 3 4
  mls qos vlan-based
  mls qos trust cos
  switchport
  switchport trunk encapsulation dot1q
  switchport mode trunk
```

Native IOS CoS-to-DSCP Mapping Configuration for Layer 2 Access Switches

Cisco follows the IETF recommendations for setting the DSCP classification values for both the VoIP control plane traffic and VoIP bearer or media plane traffic. The recommended settings are DSCP=AF31 for VoIP control plane and DSCP=EF for VoIP bearer plane. To map the Layer 2 settings correctly to these DSCP values, you must modify the default CoS-to-DSCP mappings.



Note

The Catalyst 6000 numerical values of 26 and 46 correlate to DSCP=AF31 and DSCP=EF, respectively. This is done in global configuration mode.

The command for this configuration is

```
mls qos map cos-dscp 0 8 16 26 32 46 56 0
```

Configure the QoS Policies and Layer 3 Access Lists for VoIP Control Traffic Classification

The QoS configuration for the Native Cisco IOS Catalyst 6000 is very similar to the WAN router Cisco IOS configurations, with the exception of using policing for marking traffic flows and applying service policies to VLAN interfaces. The physical gigabit Ethernet uplink ports are configured to use VLAN-based QoS with the **mls qos vlan-based** Native Cisco IOS interface commands. Finally, the **service-policy** is applied to all VLAN traffic *inbound* on the uplink.

In the following example, three classes are defined: one for the VoIP media stream, one for the control traffic, and the last for all other traffic. Traffic is filtered for these classes based on Layer 3 or 4 source and destination IP addresses and ports. Each of these classes is referenced in the **Voice-QoS** policy map. In the **policy-map** statements, a policing function is used to classify all traffic that meets the entrance criteria matched with the **class-map** access lists.



Note

The Catalyst 6000 Native Cisco IOS software does not support the **set ip dscp** commands. Instead, the policing algorithm is used for classifying traffic.

In this scenario, the policing code tags the traffic flows with DSCP values of AF31 for VoIP control traffic, EF for VoIP Media traffic, and 0 for all other packets. The size of the "8000" flows is low enough that any traffic will solicit tagging using the syntax **conform-action set-dscp-transmit 26**.

```

class-map match-all VoIP-Control
  match access-group 100
class-map match-all VoIP-RTP
  match access-group 101
class-map match-all Routine
  match access-group 102
!
!
policy-map Voice-QoS
  class VoIP-Control
    police 8000 8000 8000 conform-action set-dscp-transmit 26 exceed action transmit
  class VoIP-RTP
    police 8000 8000 8000 conform-action set-dscp-transmit 46 exceed-action transmit
  class Routine
    police 8000 8000 8000 conform-action set-dscp-transmit 0 exceed-action transmit
!
! access-list 100 looks for VoIP Control Traffic
access-list 100 permit tcp any any range 2000 2002
access-list 100 permit tcp any any eq 1720
access-list 100 permit tcp any any range 11000 11999
access-list 100 permit udp any any eq 2427
!
! access-list 101 looks for VoIP Bearer Traffic
access-list 101 permit udp any any range 16384 32767
!
! access-list 102 filters for routine traffic
access-list 102 permit ip any any
!
interface GigabitEthernet2/2
  description trunk port to layer 2-only cat4k
  no ip address
  wrr-queue cos-map 1 2 2
  wrr-queue cos-map 2 1 3 4
  ! inform the port that QoS will be VLAN-Based
  mls qos vlan-based
  switchport
  switchport trunk encapsulation dot1q
  switchport mode trunk
!

```

```

interface GigabitEthernet3/1
  description trunk port to layer 2-only 3500
  no ip address
  wrr-queue cos-map 1 2 2
  wrr-queue cos-map 2 1 3 4
  ! inform the port that QoS will be VLAN-Based
  mls qos vlan-based
  switchport
  switchport trunk encapsulation dot1q
  switchport mode trunk
!
interface Vlan111
  description voice vlan on cat4k
  ip address 10.1.111.77 255.255.255.0
  ip helper-address 10.1.10.10
  no ip redirects
  ! apply the QoS policy as an inbound policy
  service-policy input Voice-QoS
  standby 111 ip 10.1.111.1
!
interface Vlan112
  description voice vlan on 3500
  ip address 10.1.112.77 255.255.255.0
  ip helper-address 10.1.10.10
  no ip redirects
  ! apply the QoS policy as an inbound policy
  service-policy input Voice-QoS
  standby 112 ip 10.1.112.1

```

```

ios6k#sh mls qos
QoS is enabled globally
Microflow policing is enabled globally

```

QoS is vlan-based on the following interfaces:

```

Vl111 Vl112 Gi2/2 Gi3/1 Gi3/2 Gi3/3
Gi3/4 Gi3/5 Gi3/6 Gi3/7 Gi3/8 Gi4/1 Gi4/2 Gi4/3 Gi4/4 Gi4/5
Gi4/6 Gi4/7 Gi4/8 Fa9/1 Fa9/2 Fa9/3 Fa9/4 Fa9/5 Fa9/6 Fa9/7
Fa9/8 Fa9/9 Fa9/10 Fa9/11 Fa9/12 Fa9/13 Fa9/14 Fa9/15 Fa9/16 Fa9/17
Fa9/18 Fa9/19 Fa9/20 Fa9/21 Fa9/22 Fa9/23 Fa9/24 Fa9/25 Fa9/26 Fa9/27
Fa9/28 Fa9/29 Fa9/30 Fa9/31 Fa9/32 Fa9/33 Fa9/34 Fa9/35 Fa9/36 Fa9/37
Fa9/38 Fa9/39 Fa9/40 Fa9/41 Fa9/42 Fa9/43 Fa9/44 Fa9/45 Fa9/46 Fa9/47
Fa9/48

```

QoS global counters:

```
Total packets: 16750372458300
```

```
Packets dropped by policing: 55930847232
```

```
IP packets with TOS changed by policing: 16750372458300
```

```
IP packets with COS changed by policing: 55945330688
```

```
Non-IP packets with COS changed by policing: 16750372458300
```

Summary

As described in this chapter, the following general guidelines and recommendations apply when configuring a Cisco AVVID network in a campus environment:

- Multiple queues are required on all interfaces to guarantee voice quality.
- To enable fast convergence, use UplinkFast in wiring closet switches that have multiple egress queues, such as the Catalyst 2900 XL, 3500, 2948G, 2980G, 4000, and 6000 switches.
- Set all Cisco AVVID control and management traffic to maximum CoS and ToS values of 3.
- Never allow PC applications to send traffic at a CoS or ToS value of 4-7.
- Distribution layer switches must have the ability to map Layer 3 ToS to Layer 2 CoS values.



Building a Branch Office

This chapter covers design considerations and recommendations for adding a branch office to your Cisco AVVID network.

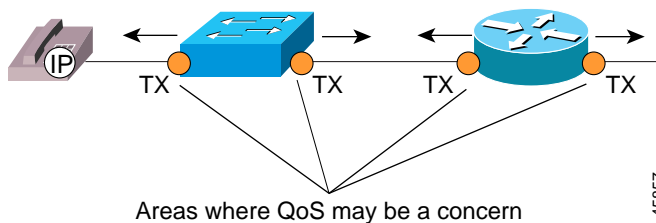
Recommended Branch Office Designs

The traditional branch office design for up to 100 users consists of a branch router and an Ethernet switch. The router handles all IP routing and WAN connectivity. The local PCs are connected to a small Ethernet switch that also connects to the router. There are two areas of concern for voice quality within the branch office:

- Voice quality across the WAN
- Voice quality within the branch office

This chapter addresses branch office design, IP addressing, and voice quality within the branch office. Details of WAN QoS tools for ensuring voice quality across the WAN are covered in Chapter 5, “Implementing a Wide Area Network.”

Figure 4-1 shows areas where QoS concerns might arise in a branch office.

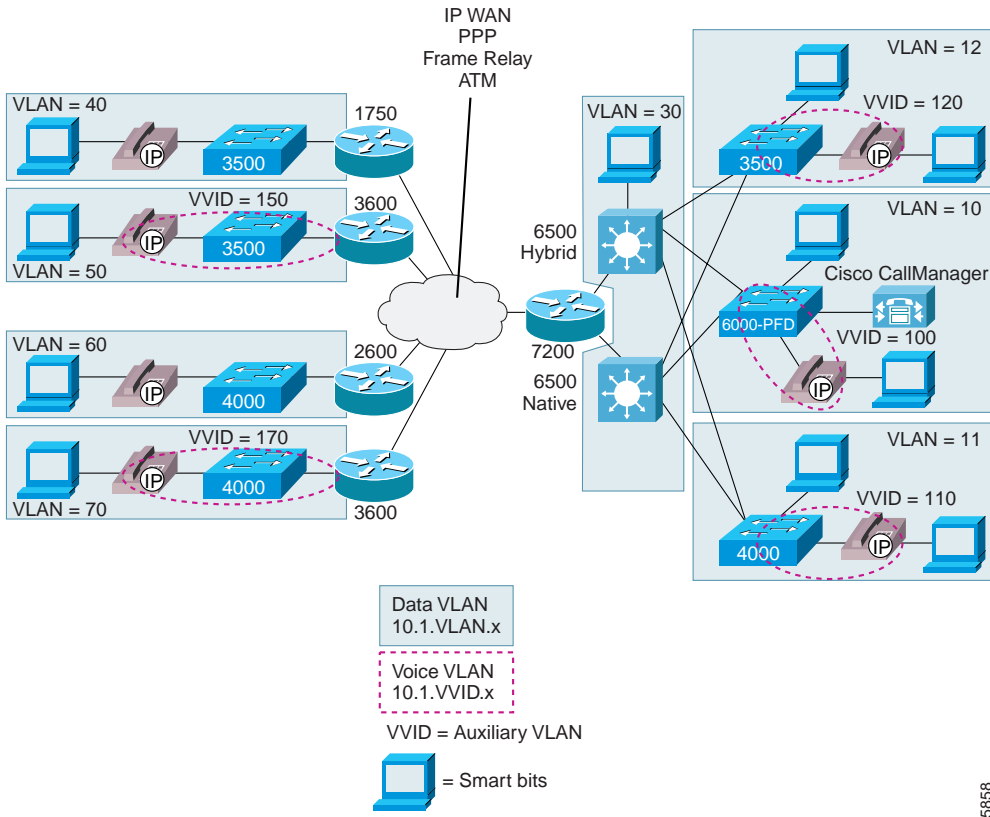
Figure 4-1 QoS Considerations for a Branch Office Connection

Typically, when branch offices are designed, only a single IP subnet is used for each office. Changing this configuration is seldom feasible because doing so affects the enterprise-wide routing scheme. Therefore, realistic branch office designs must examine three IP addressing options for IP phones:

- Configure two VLANs by dividing the existing remote office IP address space into subnets, and use 802.1Q for trunking if the router supports trunking.
- Configure two VLANs by dividing the existing remote office IP address space into subnets, and use secondary IP addressing on the Ethernet interface of the router.
- Use a single IP address space and VLAN at each remote office.

Each scenario described in this chapter uses a single cable to install an IP phone in the branch office because this is the more common methodology in practice today. Figure 4-2 shows the general model for the branch office configurations described in this chapter.

Figure 4-2 General Model for Branch Office Connections



45858

Using 802.1Q for Trunking Separate Voice and Data Subnets at the Branch Office

You should always use separate VLANs for voice and data when there is an option to segment the existing IP address space of the branch office. Ethernet switches that support only Layer 2 services, such as the current Catalyst 3500 and 4000 series, are used in almost every branch office design. When this is the case, the branch WAN routers trunk the separate VLANs from the Ethernet switch. This is done using 802.1Q trunking on the router and switch.

As has been described throughout this guide, User Priority bits in the 802.1p portion of the 802.1Q standard header are used to provide prioritization in Ethernet switches. This is a vital component in designing Cisco AVVID networks. Cisco IOS Release 12.1(5)T includes the Modular CLI QoS code. These additions enable the mapping of Layer 3 tagged VoIP packets coming from the WAN to be classified with the appropriate Layer 2 802.1D User Priority marking for proper queuing on the branch office Catalyst Ethernet switch.

At the headquarters, the Catalyst 6000 correlates all Layer 3 ToS settings to the correct Layer 2 CoS value at the ingress interface.

Catalyst 3600 Branch Office Router Using 802.1Q Trunking

The following Catalyst 3600 configuration includes the Layer 3 to Layer 2 classification mappings for incoming voice packets:

```
class-map L3-to-L2_VoIP-RTP
  match ip dscp ef
class-map L3-to-L2_VoIP-Cntrl
  match ip dscp af31
policy-map output-L3-to-L2
  class L3-to-L2_VoIP-RTP
    set cos 5
  class L3-to-L2_VoIP-Cntrl
    set cos 3
!
interface FastEthernet1/0
  description Catalyst 3500 Branch Office Switch
  no ip address
  ip route cache policy
  no ip mroute-cache
  load-interval 30
  speed 100
  full-duplex
!
interface FastEthernet1/0.50
  description native subnet 10.1.50.0 data
  encapsulation dot1Q 50
  ip address 10.1.50.1 255.255.255.0
  service-policy output output-L3-to-L2
  ip route cache policy
  no ip mroute-cache
!
interface FastEthernet1/0.150
  description native subnet 10.1.150.0 voice
  encapsulation dot1Q 150
  ip address 10.1.150.1 255.255.255.0
  ip helper-address 10.1.10.10
  service-policy output output-L3-to-L2
  ip route cache policy
  no ip mroute-cache
```

Catalyst 4000 Using 802.1Q Trunking

The following example shows a branch office configuration for a Catalyst 4000 using 802.1Q trunking:

```
cat4k> (enable) set vlan 70 name data70
cat4k> (enable) set vlan 170 name voice170
cat4k> (enable) set vlan 70 2/1-48
cat4k> (enable) set port host 2/1-48
cat4k> (enable) set port auxiliaryvlan 2/1-48 170
cat4k> (enable) set port speed 2/1-49 100
cat4k> (enable) set port duplex 2/1-49 full
cat4k> (enable) set trunk 2/49 on dot1q 1-1005
```

Catalyst 3500 Using 802.1Q Trunking

The following example shows a branch office configuration for a Catalyst 3500 using 802.1Q trunking:

```
interface FastEthernet0/1
  description DOT1Q port to IP Phone
  switchport trunk encapsulation dot1q
  switchport trunk native vlan 50
  switchport mode trunk
  switchport voice vlan 150
  spanning-tree portfast
!
interface FastEthernet0/15
  description Port to 3640 (supports Dot1q)
  duplex full
  speed 100
  switchport trunk encapsulation dot1q
  switchport trunk native vlan 50
  switchport trunk allowed vlan 1,50,150
  switchport mode trunk
```

Using Secondary IP Addressing for Separate Voice and Data Subnets at the Branch Office

Separate VLANs for voice and data are still recommended in cases where the branch router does not support 802.1Q trunking. For example, the Cisco 1750 router does not support trunking, but a logical separation of voice and data traffic is still desirable. An alternative to trunking is to use secondary IP addressing on the Cisco router. The following example shows this type of configuration for the Cisco 1750 at a branch office:

```
interface FastEthernet0
  description to Catalyst 3500
  mac-address 0000.1750.0001
  ip address 10.1.40.1 255.255.255.128
  ip address 10.1.40.129 255.255.255.128 secondary
  ip helper-address 10.1.10.10
  no ip mroute-cache
  speed 100
  full-duplex
```

Classifying VoIP Control Traffic at the Branch Office

The remote branch router also must classify VoIP Control traffic leaving the local subnets and heading for a Cisco CallManager or VoIP gateway across the WAN. You can achieve this classification of VoIP Control traffic through the use of policy-based routing and route maps on the ingress Ethernet interface.

**Note**

Beginning with Release 3.0(5), Cisco CallManager includes the ability to configure the CoS and ToS values for all VoIP control and management traffic from Cisco CallManager, the IP phones, and the Skinny Protocol gateways (this does not include the AT and AS model analog gateways). With this user-configurable classification, network element access lists are no longer required for marking Skinny Protocol VoIP control traffic. H.323 and Media Gateway Control Protocol (MGCP) traffic still require external, network element marking for several more months.

The following example shows a configuration for classifying VoIP Control traffic at the branch office:

```
interface FastEthernet0
    mac-address 0000.1750.0001
    ip address 10.1.60.1 255.255.255.0
    ip helper-address 10.1.10.10
    ! Attach the route-map to the FastEthernet interface
    ip policy route-map Set-IP-QoS
    no ip mroute-cache
    load-interval 30
    speed auto
    full-duplex

!
! Match all Skinny, H.323 and MGCP Control Traffic
! Skinny marking is required only until
! Cisco CallManager Release 3.0(5).
access-list 101 permit tcp any any range 2000 2002
access-list 101 permit tcp any any eq 1720
access-list 101 permit tcp any any range 11000 11999
access-list 101 permit udp any any eq 2427
!
! Match all VoIP RTP Traffic
access-list 102 permit udp any any range 16384 32767
!
! Match all other traffic
access-list 103 permit ip any any
!
! Set all Skinny, H.323 and MGCP Control traffic, matched
! with ac 101 to IP Precedence 3
route-map Set-IP-QoS permit 10
    match ip address 101
    set ip precedence flash
!
! Just match VoIP RTP Traffic. Do not change the
! default classification of ToS=5
route-map Set-IP-QoS permit 20
    match ip address 102
!
! Make sure all data traffic is set to IP Precedence 0
route-map Set-IP-QoS permit 30
    match ip address 103
    set ip precedence routine
```

Using a Single Subnet at the Branch Office

It might be necessary to use a single IP address space for branch offices in situations where it is impractical either to allocate an additional IP subnet for IP phones or to divide the existing IP address space into an additional subnet at the remote branch. When this is the case, there is still a need to prioritize voice above data at both Layer 2 and Layer 3.

As described in Chapter 2, “Connecting IP Phones,” Layer 3 classification is already taken care of because the phone sets the Type of Service (ToS) bits in all media streams to an IP Precedence value of 5. (With Cisco CallManager Release 3.0(5), this marking changed to a Differentiated Services Code Point (DSCP) value of EF.) However, to ensure that there is Layer 2 classification for admission to the multiple queues in the branch office switches, the phone must also use the User Priority bits in the Layer 2 802.1p header to provide Class of Service (CoS) marking. This can be done only by having the switch look for 802.1p headers on the native VLAN. The two Ethernet switches employed most often in branch office designs, the Catalyst 3500 and 4000, use different configuration commands to accomplish this, as described in the following sections.

Cisco 1750 Single Subnet Configuration

The Cisco 1750 series router does not support either Inter-Switch Link (ISL) or 802.1Q Ethernet trunking. The following example shows a single subnet configuration for the Cisco 1750 router:

```
interface FastEthernet0
  mac-address 0000.1750.0001
  ip address 10.1.40.1 255.255.255.0
  ip helper-address 10.1.10.10
  ip policy route-map Set-IP-QoS
  no ip mroute-cache
  load-interval 30
  speed auto
  full-duplex
```

Catalyst 3500 Single Subnet Configuration

The Catalyst 3500 supports the use of an 802.1p-only option when configuring the auxiliary VLAN. This allows the IP phone to tag VoIP packets with a CoS of 5 on the native VLAN, while all PC data traffic is sent untagged. The following example shows a single subnet configuration for the Catalyst 3500:

```
interface FastEthernet0/2
description Port to IP Phone in single subnet
  switchport trunk encapsulation dot1q
  switchport trunk native vlan 40
  switchport mode trunk
  switchport voice vlan dot1p
  spanning-tree portfast
!
interface FastEthernet0/15
description Port to 1750 Router in single subnet
load-interval 30
duplex full
speed 100
switchport access vlan 40
```

Catalyst 2600 Single Subnet (no Trunking) Configuration

The following example shows a single subnet configuration for the Catalyst 2600:

```
interface FastEthernet1/0
mac-address 0000.2600.0001
ip address 10.1.60.1 255.255.255.0
ip helper-address 10.1.10.10
ip policy route-map Set-IP-QoS
no ip mroute-cache
load-interval 30
speed 100
full-duplex
```

Catalyst 4000 Single Subnet Configuration

Unlike the Catalyst 3500 and 6000, the Catalyst 4000 does not support a dot1p-only option for configuring the auxiliary VLAN. As a workaround, you should configure the IP phones connected to the Catalyst 4000, so that the auxiliary VLAN ID matches the Port VLAN ID (PVID) or native VLAN ID. This ensures that the phone can still send its packets tagged with a CoS of 5. The following example shows this configuration for the Catalyst 4000:

```
cat4k> (enable) set vlan 60 name 171.69.60.0_data
cat4k> (enable) set vlan 60 2/1-49
cat4k> (enable) set port host 2/1-49
cat4k> (enable) set port auxiliaryvlan 2/1-48 60
```

Summary

As described in this chapter, the following general guidelines and recommendations apply when configuring a branch office on a Cisco AVVID network:

- The branch WAN router must support the advanced QoS tools for Cisco AVVID networks.
- Use a switch that supports multiple queues.
- Currently, there is no way to pass Layer 3 ToS classification to Layer 2 CoS in the routers. However, future releases of Cisco IOS will contain additional QoS features that allow for these mappings.



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."

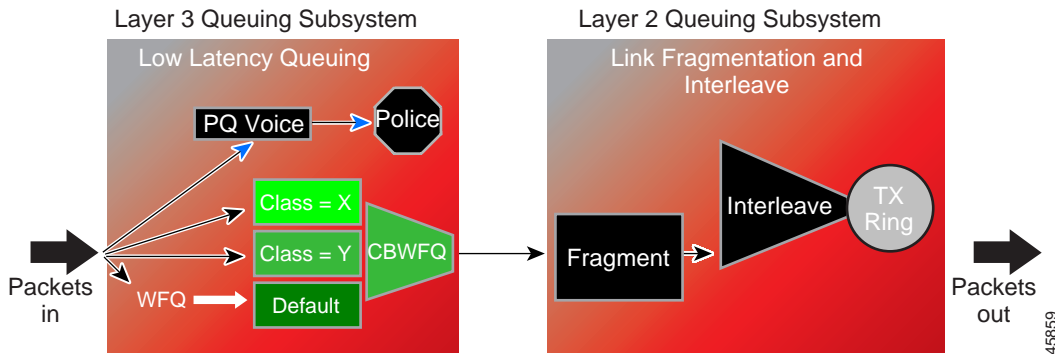
Queuing

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.

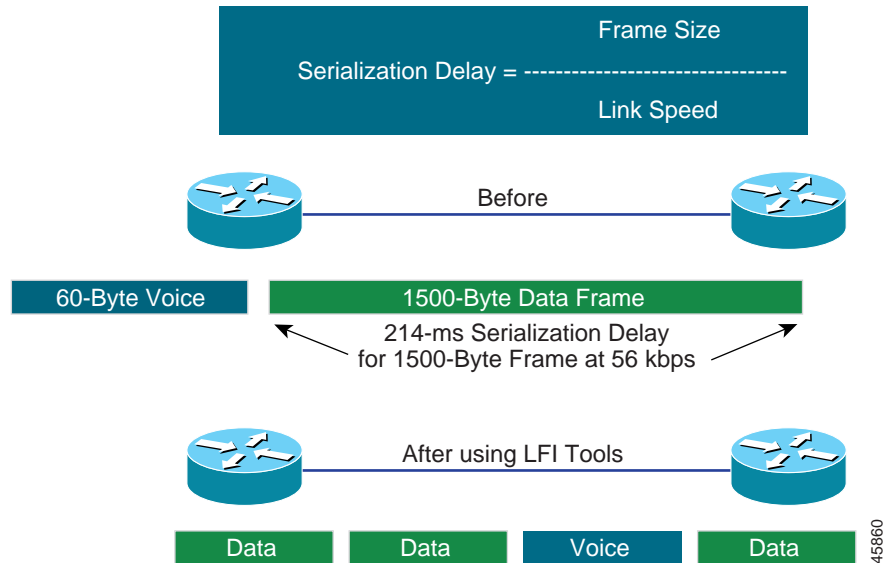
Link Fragmentation and Interleaving

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

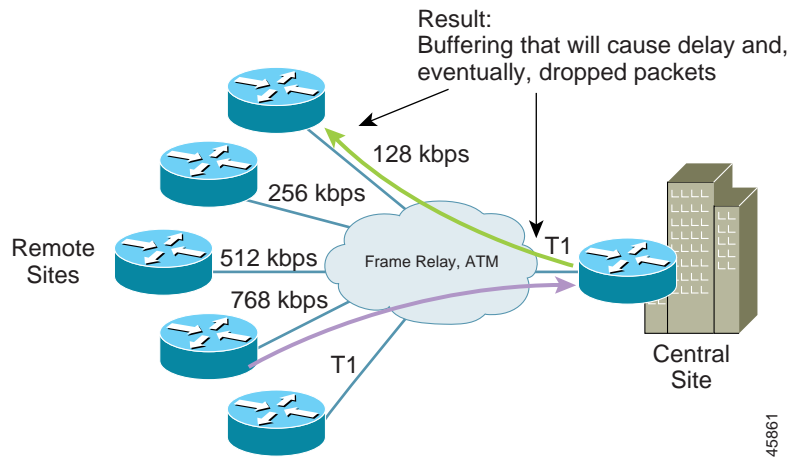
**Note**

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.

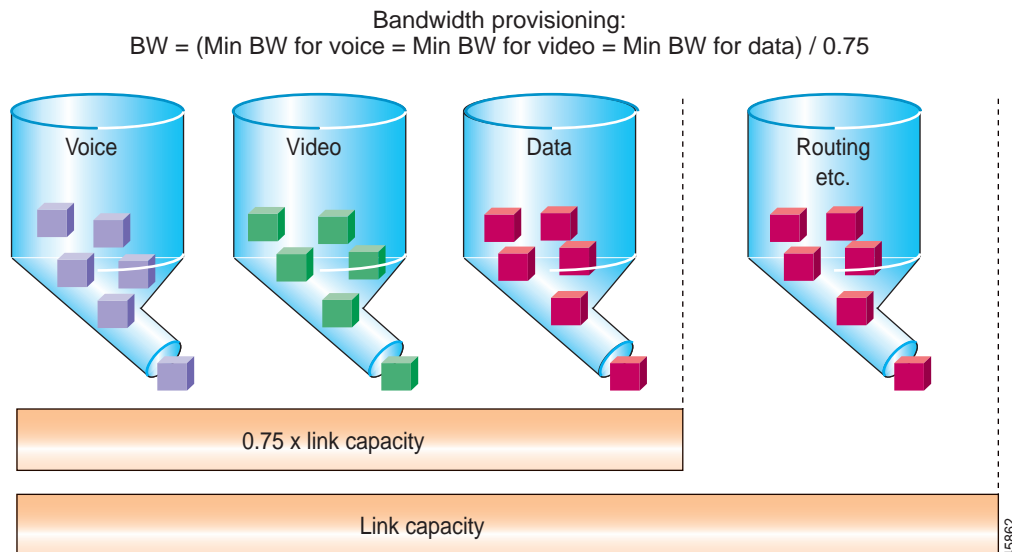
Traffic leaving a high-speed interface such as a T1 line at a central site often terminates at a remote site that may have a much slower link speed (for example, 56 kbps). This is quite common and, in fact, has been one of the big selling points for Frame Relay. In Figure 5-3, the T1 interface on the 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

Network Provisioning

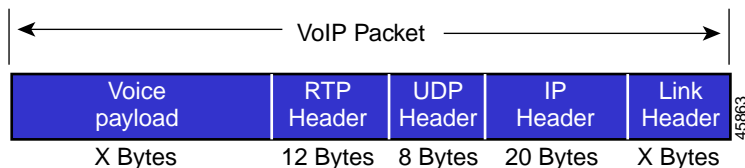
Properly provisioning the network bandwidth is a major component of designing a successful Cisco AVVID network. You can calculate the required bandwidth by adding the bandwidth requirements for each major application (for example, voice, video, and data). This sum then represents the minimum bandwidth 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

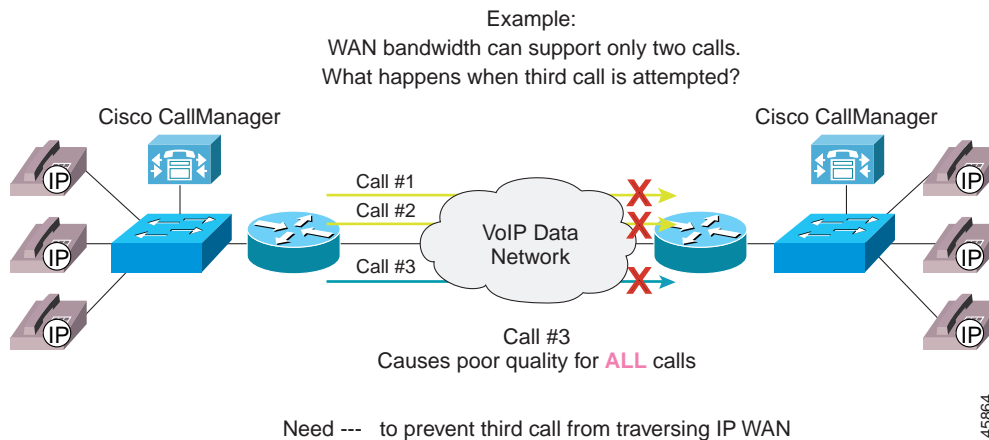
Call Admission Control

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.



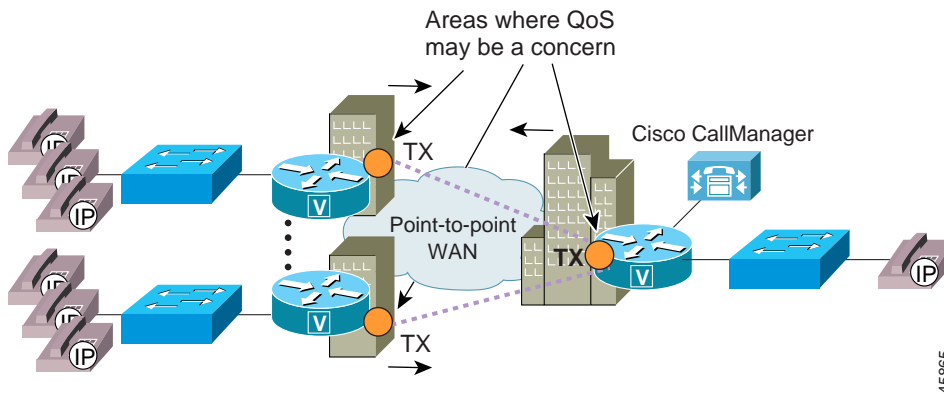
Note

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: 10.1.60.98, destination: 10.1.10.98, id: 0x0322, ttl: 63,
TOS: 0 prot: 17, source port 1024, destination port 7
```

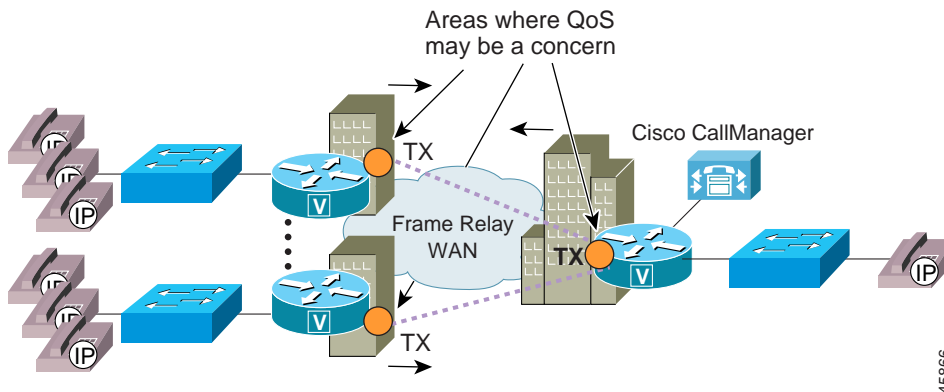
```
1750# sh policy interface multilink1
Multilink1
output : QoS-Policy-256k
Class VoIP-RTP
  Weighted Fair Queueing
    Strict Priority
    Output Queue: Conversation 264
      Bandwidth 100 (kbps)
      (pkts matched/bytes matched) 28100/5675882
      (pkts discards/bytes discards) 0/0
Class VoIP-Control
  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
```

```

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

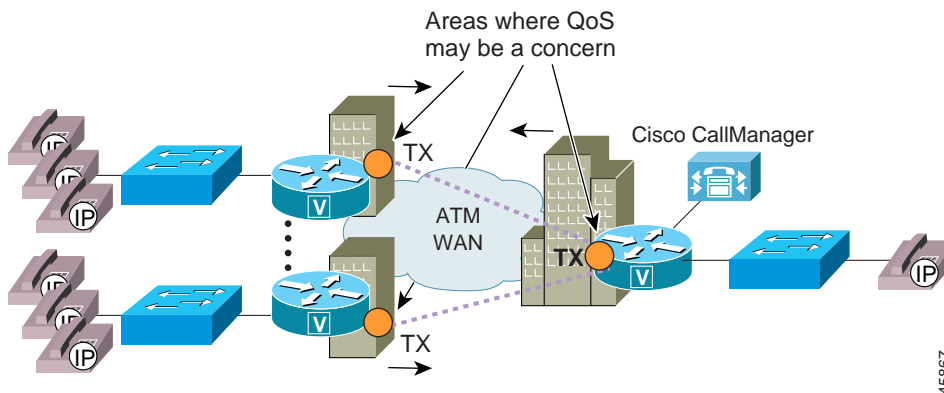
shaping inactive
traffic shaping drops 0

```

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.

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.



Numerics

- 10/100 ports **3-28**
- 1750 router **4-9**
- 1P1Q4T configuration **3-13**
- 1P2Q2T configuration **3-14**
- 1Q4T configuration **3-13**
- 1Q-FIFO configuration **3-28**
- 2Q1T configuration **3-24, 3-28**
- 2Q2T configuration **3-14**
- 4Q1T configuration **3-27**
- 7200 WAN router **3-37**
- 802.1Q access trunks
 - at branch office **4-4**
 - on Catalyst 2900 XL and 3600 XL **2-7**
 - on Catalyst 3500 **4-6**
 - on Catalyst 3600 **4-5**
 - on Catalyst 4000 **4-6**
 - on Catalyst 4000 and 6000 **2-6**
- 8Q-FIFO configuration **3-28**

A

- access control list (ACL) **3-36**
- access layer switch
 - Catalyst 3500 **3-27**
 - Catalyst 4000 **3-23**
 - Catalyst 6000 **3-11**
 - for IP phones **2-16**
- access trunks
 - on Catalyst 2900 XL and 3500 XL **2-7**
 - on Catalyst 4000 and 6000 **2-6**
- ACL **3-36, 3-43**
- adaptive jitter buffer **1-4**
- additional information **xv**
- addresses
 - for IP phones **2-5, 2-12, 2-17**
 - for SoftPhone **2-15**
 - secondary IP addressing at branch office **4-7**
- admission control **5-10**
- assistance with Cisco products **xvi**
- Asynchronous Transfer Mode (ATM) **5-30**
- ATM **5-30, 5-40**
- ATM-to-Frame-Relay **5-35**
- audience for this guide **xii**

B

bandwidth

- consumption **5-8**
- provisioning **1-7, 5-7**

Bc **5-23**Be **5-23**

branch office

- design considerations **4-1**
- Frame-Relay **5-41**
- traffic classification **4-7**

buffer

- jitter **1-4**

burst rate **5-23**

Ccall admission control **5-10**

campus

- ATM configuration **5-40**
- design considerations **3-1**

Catalyst 2600 **4-10**

Catalyst 2900 XL

- 802.1Q access trunk configuration **2-7**
- port configuration for IP phones **2-5, 2-14**
- trust boundary extension **2-10**

Catalyst 2948G **2-10**Catalyst 2980G **2-10**

Catalyst 3500

- 802.1Q access trunk configuration **4-6**
- access layer switching **3-27**
- port scheduling **3-28**
- queuing schemes **3-28**
- receive interface **3-28**
- single subnet configuration **4-10**

Catalyst 3500 XL

- 802.1Q access trunk configuration **2-7**
- port configuration for IP phones **2-5, 2-14**
- trust boundary extension **2-10**

Catalyst 3600 **4-5**

Catalyst 4000

- 802.1Q access trunk configuration **2-6**
- 802.1Q trunk configuration **4-6**
- access layer switching **3-23**
- port configuration for IP phones **2-4, 2-13**
- port scheduling **3-23**
- QoS **3-25**
- queuing schemes **3-23**
- receive interface **3-23**
- single subnet configuration **4-11**
- transmit interface **3-24**
- trust boundary extension **2-10**

Catalyst 6000

- 802.1Q access trunk configuration **2-6**
- access layer switching **3-11**
- distribution layer switching **3-31**
- Native IOS **3-38**

port configuration for IP phones **2-4, 2-12**
 port scheduling **3-13**
 queuing schemes **3-13**
 receive interface **3-13**
 transmit interface **3-14**
 transmit queue **3-21**
 trust boundary extension **2-9**
 CCO **xvi**
 CIR **5-22, 5-24**
 Cisco Connection Online (CCO) **xvi**
 Cisco IP Phones **2-1**
 classification
 for IP phones **2-7, 2-12, 2-15, 2-17**
 of traffic **1-5, 3-6, 4-7, 5-2**
 class of service (CoS) **1-5**
 CODEC **5-14**
 comments on this guide **xviii**
 committed burst rate (Bc) **5-23**
 committed information rate (CIR) **5-22**
 compressed RTP (cRTP) **5-14**
 compressed voice **5-14**
 control traffic **3-6, 3-32, 3-40, 4-7, 5-11**
 conventions used in this guide **xiii**
 CoS **1-5, 3-35, 3-42**
 CoS-to-DSCP mappings **3-22, 3-35, 3-43**
 cRTP **5-14, 5-18, 5-26, 5-33, 5-41**

D

delay
 of packets **1-2**
 serialization **1-3, 5-4**
 distribution layer switch **3-31**
 documentation
 additional **xv**
 CD-ROM **xv**
 obtaining **xv**
 ordering from Cisco **xvi**
 World Wide Web **xv**
 your feedback **xviii**
 DSCP trust values **3-33, 3-40**
 duplex settings
 IP phones **2-3, 2-11, 2-16**
 SoftPhone **2-15**

E

excess burst rate (Be) **5-23**

F

FIFO configuration **3-23**
 Frame-Relay **5-21, 5-41**
 Frame-Relay-to-ATM **5-35**
 FRF.12 **5-25**
 FRF.8 **5-35**

G

gigabit Ethernet ports **3-28**

H

H.323 protocol **3-9**

help with Cisco products **xvi**

I

IP addressing **2-5, 2-12, 2-15, 2-17, 4-7**

IP phones

connecting **2-1**

IP addressing **2-5, 2-12, 2-17**

multiple-cable connection **2-11**

on a separate switch **2-16**

port configuration on Catalyst 2900 XL and
3500 XL **2-5**

port configuration on Catalyst 4000 and
6000 **2-4**

QoS issues **2-1**

queuing **2-7, 2-12, 2-15, 2-17, 3-16, 3-26, 3-30**

single-cable connection **2-2**

speed and duplex settings **2-3, 2-11, 2-16**

traffic classification **2-7, 2-12, 2-15, 2-17**

J

jitter

adaptive buffer **1-4**

described **1-2**

L

Layer 2 switch **3-34, 3-42**

LFI

described **5-4**

for internetworking WANs **5-37**

for point-to-point WAN **5-17**

link fragmentation and interleaving (LFI) **5-4**

LLQ

described **5-2**

for ATM **5-34**

for Frame Relay **5-26**

for internetworking **5-41**

for WAN configurations **5-18**

low-latency queuing (LLQ) **5-2**

M

management traffic **3-6**

mapping to DSCP values **3-22, 3-34, 3-35, 3-41,
3-43**

Media Gateway Control Protocol
(MGCP) **3-10**

MGCP **3-10**

minimum CIR **5-24**

MLS **3-21**

model of VoIP network **1-8**

MSFC **3-21**

Multilink Point-to-Point Protocol (MLP) **5-17**

multiple-cable connection of IP phones **2-11**

N

Native IOS **3-38**

network

congestion **1-2**

general VoIP model **1-8**

provisioning **1-7**

provisioning bandwidth **5-7**

quality **1-2**

notational conventions **xiii**

O

organization of this guide **xii**

P

packet

classification **1-5**

phones

connecting **2-1**

IP addressing **2-5, 2-12, 2-17**

multiple-cable connection **2-11**

on separate switch **2-16**

port configuration on Catalyst 2900 XL and 3500 XL **2-5**

port configuration on Catalyst 4000 and 6000 **2-4**

QoS issues **2-1**

queuing **2-7, 2-12, 2-15, 2-17, 3-16, 3-26, 3-30**

single-cable connection **2-2**

speed and duplex settings **2-3, 2-11, 2-16**

traffic classification **2-7, 2-12, 2-15, 2-17**

point-to-point WAN **5-16**

port configuration for IP phones

on Catalyst 2900 XL and 3500 XL **2-5**

on Catalyst 4000 and 6000 **2-4**

port scheduling

Catalyst 3500 **3-28**

Catalyst 4000 **3-23**

Catalyst 6000 **3-13**

protocols

H.323 **3-9**

MGCP **3-10**

Skinny Gateway **3-8**

Skinny Station **3-8**

provisioning network bandwidth **1-7, 5-7**

purpose of this guide **xi**

PVC **5-32**

Q

QoS

- access layer parameters **3-15**
 - branch office design **4-1**
 - campus design **3-1**
 - Catalyst 4000 **3-25**
 - configuration **3-43**
 - general principles for VoIP **1-9**
 - IP phones **2-1**
 - issues **1-1**
 - Native IOS on Catalyst 6000 **3-39**
 - overview **1-1**
 - tools **1-5, 5-11**
 - WAN implementation **5-1**
 - why needed **1-1**
- quality of service (QoS) **1-1**
- queuing
- Catalyst 3500 **3-28**
 - Catalyst 4000 **3-23**
 - Catalyst 6000 **3-13**
 - for IP phones **2-7, 2-12, 2-15, 2-17, 3-16, 3-26, 3-30**
 - for WANs **5-2**
 - number of queues **3-5**
 - overview **1-7**
 - scheduling algorithms **3-4**
 - VoIP control traffic **3-32, 3-40**

R

receive interface configuration

- Catalyst 3500 **3-28**
- Catalyst 4000 **3-23**
- Catalyst 6000 **3-13**

S

scheduling

- ports **3-13, 3-23, 3-28**
- separate subnets for voice and data **4-4, 4-7**
- serialization delay **1-3, 5-4**
- shaping traffic **5-6, 5-22**
- show command **3-17, 3-22, 3-26**
- single-cable connection of IP phones **2-2**
- single subnet at branch office **4-9**
- Skinnny Protocol **3-8**
- SoftPhone **2-15**
- speed and duplex settings
 - IP phones **2-3, 2-11, 2-16**
 - SoftPhone **2-15**
- subnets
 - secondary IP addressing **4-7**
 - separate voice and data at branch office **4-4**
 - single subnet at branch office **4-9**

T

TAC **xvi, xvii**
 technical assistance **xvi**
 tools for QoS **1-5, 5-11**
 ToS **1-5**
 ToS-to-DSCP mappings **3-22, 3-34, 3-41**
 traffic
 access control lists **3-36, 3-43**
 classification **3-6, 5-2**
 marking **3-6**
 shaping **5-6, 5-22**
 transmit interface configuration
 10/100 ports **3-28**
 Catalyst 4000 **3-24**
 Catalyst 6000 **3-14**
 gigabit Ethernet ports **3-28**
 transmit queue
 Catalyst 6000 **3-21**
 configuration **3-32, 3-40**
 trust boundary extension
 Catalyst 2900 XL **2-10**
 Catalyst 2948G **2-10**
 Catalyst 2980G **2-10**
 Catalyst 3500 XL **2-10**
 Catalyst 4000 **2-10**
 Catalyst 6000 **2-9**
 trust DSCP **3-33, 3-40**

TX-Ring **5-12**
 type of service (ToS) **1-5**

U

uplink interface **3-21, 3-26, 3-30**

V

VAD **5-15**
 verifying configurations **3-17, 3-22, 3-26, 5-20, 5-28**
 voice activity detection (VAD) **5-15**
 voice compression **5-14**
 VoIP
 control traffic **3-32, 3-40, 4-7, 5-11**
 general network model **1-8**
 QoS principles **1-9**

W

WAN
 7200 router **3-37**
 ATM **5-30**
 Frame-Relay **5-21**
 Frame-Relay-to-ATM **5-35**
 FRF.12 **5-25**
 FRF.8 **5-35**
 implementation **5-1**

- internetworking **5-35**
- LFI **5-17, 5-37**
- LLQ **5-18, 5-34**
- MLP **5-17**
- point-to-point **5-16**
- PVC **5-32**
- QoS issues **5-1**
- QoS tools **5-11**
- wiring closet
 - Catalyst 3500 **3-27**
 - Catalyst 4000 **3-23**
 - Catalyst 6000 **3-11**