SYNGRESS®

1 YEAR UPGRADE
BUYER PROTECTION PLAN

Cisco® AVVID and IP Telephony

Design & Implementation

Design and Deploy IP Telephony Solutions

- Step-by-Step Instructions for Using AVVID Applications in Single Site and Multiple Site Solutions
- Complete Coverage of Voice and Video Gatekeeper Design
- Hundreds of Configuring & Implementing and Designing & Planning Sidebars, FAQs, and Case Studies!

Robert Padjen
Larry Keefer
Sean Thurston
Jeff Bankston
Michael E. Flannagan
Martin Walshaw Technical Editor

callisma

solutions@syngress.com

With more than 1,500,000 copies of our MCSE, MCSD, CompTIA, and Cisco study guides in print, we continue to look for ways we can better serve the information needs of our readers. One way we do that is by listening.

Readers like yourself have been telling us they want an Internet-based service that would extend and enhance the value of our books. Based on reader feedback and our own strategic plan, we have created a Web site that we hope will exceed your expectations.

Solutions@syngress.com is an interactive treasure trove of useful information focusing on our book topics and related technologies. The site offers the following features:

- One-year warranty against content obsolescence due to vendor product upgrades. You can access online updates for any affected chapters.
- "Ask the Author" customer query forms that enable you to post questions to our authors and editors.
- Exclusive monthly mailings in which our experts provide answers to reader queries and clear explanations of complex material.
- Regularly updated links to sites specially selected by our editors for readers desiring additional reliable information on key topics.

Best of all, the book you're now holding is your key to this amazing site. Just go to **www.syngress.com/solutions**, and keep this book handy when you register to verify your purchase.

Thank you for giving us the opportunity to serve your needs. And be sure to let us know if there's anything else we can do to help you get the maximum value from your investment. We're listening.

www.syngress.com/solutions

SYNGRESS®

SYNGRESS®

1 YEAR UPGRADE
BUYER PROTECTION PLAN

Cicso® AVVID and IP Telephony

Design & Implementation

Robert Padjen
Larry Keefer
Sean Thurston
Jeff Bankston
Michael E. Flannagan
Martin Walshaw Technical Editor

KEY	SERIAL NUMBER
001	9KPARFAHFV
002	L2DVNLA4UT
003	4ASBNL56AS
004	G7YAKETP39
005	8HJDLRG96U
006	Z64SH5Y89W
007	33RPWRJKL6
008	FV7BRD25GS
009	B8X25GVAST
010	WE4VG9LWL4

PUBLISHED BY
Syngress Publishing, Inc.
800 Hingham Street
Rockland, MA 02370

Cisco® AVVID and IP Telephony Design & Implementation

Printed in the United States of America

1 2 3 4 5 6 7 8 9 0

ISBN: 1-928994-83-0

Technical Editor: Martin Walshaw
Technical Reviewer: Sean Thurston
Co-Publisher: Richard Kristof
Acquisitions Editor: Catherine B. Nolan
Developmental Editor: Kate Glennon
Freelance Editorial Manager: Maribeth Corona-Evans
Cover Designer: Michael Kavish
Page Layout and Art by: Shannon Tozier
Copy Editor: Michael McGee
Indexer: Jennifer Coker

Distributed by Publishers Group West in the United States and Jaguar Book Group in Canada.

Acknowledgments

We would like to acknowledge the following people for their kindness and support in making this book possible.

Richard Kristof and Duncan Anderson of Global Knowledge, for their generous access to the IT industry's best courses, instructors, and training facilities.

Ralph Troupe, Rhonda St. John, Emlyn Rhodes, and the team at Callisma for their invaluable insight into the challenges of designing, deploying, and supporting world-class enterprise networks.

Karen Cross, Lance Tilford, Meaghan Cunningham, Kim Wylie, Harry Kirchner, Kevin Votel, Kent Anderson, and Frida Yara of Publishers Group West for sharing their incredible marketing experience and expertise.

Mary Ging, Caroline Hird, Simon Beale, Caroline Wheeler, Victoria Fuller, Jonathan Bunkell, and Klaus Beran of Harcourt International for making certain that our vision remains worldwide in scope.

Anneke Baeten and Annabel Dent of Harcourt Australia for all their help.

David Buckland, Wendi Wong, Daniel Loh, Marie Chieng, Lucy Chong, Leslie Lim, Audrey Gan, and Joseph Chan of Transquest Publishers for the enthusiasm with which they receive our books.

Kwon Sung June at Acorn Publishing for his support.

Ethan Atkin at Cranbury International for his help in expanding the Syngress program.

Contributors

Robert Padjen (CCNP-Security, CCDP) is a Senior Consultant with Callisma and he has written a number of texts on Cisco networking. In addition to instructing, Robert works as an expert witness in the computer and networking fields.

Mark Edwards (CCIE #7103, CCDP, CCNP, MCSE, CNE) is a Director of and Senior Network Consultant for Capital Network Solutions Ltd., based in South Wales, UK. Capital Network Solutions is a Cisco Premier Partner, specializing in Voice Access and Wireless solutions, and has completed a number of major projects for large international organizations. Their Web site can be found at www.capitalnetworks.co.uk. Mark graduated from the University of Glamorgan with a BSc (Hons) in Computer Science in 1994 and has been working as a networking consultant ever since. He now lives in Cardiff with his wife Sarah and son Cameron.

Michael E. Flannagan (CCIE #7651, CCDP, CCNA, 3COM-CSA) is Network Consulting Engineer and Team Leader in the Network Supported Accounts (NSA) Group at Cisco Systems. Mike is a member of the global Quality of Service (QoS) Team and has extensive network design experience, with emphasis on Routing Protocol design and Quality of Service mechanisms. Mike's experience, prior to joining Cisco Systems, includes enterprise network architecture, IT management, and consulting. Mike's QoS testing and research was used to recommend the implementation of various QoS mechanisms for one of the world's largest pharmaceutical companies, and he has participated in large-scale QoS designs for several major US companies. In addition to holding various certifications from Cisco, 3Com, and Nortel Networks, Mike has passed both the CCIE Routing/Switching and the CCIE Design written exams and is currently preparing for his CCIE Lab exams. He lives in Morrisville, NC.

Rob Webber (CCIE #6922) is a Senior Network Consultant with Callisma in Wakefield, MA. He has over 14 years of experience in the data networking industry, the last four as a consultant. He specializes in the design and implementation of complex networks in the financial, medical, manufacturing, and service provider industries. His expertise includes routing, switching, and security equipment from Cisco Systems and Nortel Networks. Rob holds a Bachelor of Science degree from the University of New Hampshire.

Jeff Bankston (CCDP, CCNP-Voice and Security) is the Senior Network Architect at BCI Associates. He has designed, developed, and implemented networks ranging from 75 users to enterprises in excess of 47,000 users spanning 40 buildings in a campus, WAN, and metropolitan networks. He has troubleshot the same size networks, configured and modified LAN equipment from Cisco, 3Com, Cabletron, Bay Networks, and many smaller vendors. He serves as the assistant to the Branch Technical Manager for new business development with ATM, Voice over IP, enterprise LAN/WAN development, and other such technologies. Jeff has published three books on networking, published numerous technical whitepapers and articles, and continues to write for four major technical journals. He enjoys technical testing and evaluation of vendor products in his off time, which helps him to recommend proper technologies for e-commerce environments. He enjoys teaching networking classes for Element K online distance learning (www.elementk.com) where he also develops new courses for the system. Jeff holds five major industry certifications including Cisco CCDP, CCNP Voice Specialist, and the CCNP Security Specialist. He is a Cisco CCIE candidate focusing on wireless networking.

Larry Keefer (CCNP-Voice and Security, CCDP, CIPT, BCFP, BCSD, MCSE, MCP+I, Master CAN, HP Start) is a Consultant with Callisma. His areas of specialization include design, integration, implementation, and documentation of multiple protocol and layer networks with voice, video, and data. He recently designed and implemented a multisite AVVID network utilizing CallManager 3.0x IP-PBX, IP phones, inline power switches, voice analog, and digital gateways. Prior to Callisma, Larry was a

Senior Network Engineer and Team Leader at Rush Creek Solutions. He holds a Bachelor of Science in Business Information Systems and Business Administration from Illinois State University. He has completed course work toward an M.S. degree in Computer Information Systems, University of Phoenix.

Eric Knipp (CCNP, CCDP, CCNA, CCDA, MCSE, MCP+I) is a Consultant with Callisma. He is currently engaged in a broadband optimization project for a major U.S. backbone service provider. He specializes in Cisco routers, LAN switches, Cisco's optical networking product offering as well as Microsoft NT, and network design and implementation. Eric's background includes positions as a Project Manager for a major international law firm and as a Project Manager for Nortel.

Technical Editor and Contributor

Martin Walshaw (CCIE #5629,CCNP, CCDP) is a Systems Engineer working for Cisco Systems in South Africa. His areas of specialty include IP Telephony (including all voice and video applications such as IPCC) and security, both of which keep him busy night and day. During the last 14 years, Martin has dabbled in many aspects of the IT industry, ranging from programming in RPG III and Cobol to PC sales. When Martin is not working, he likes to spend time with his expectant wife Val and his son Joshua. Without their patience, understanding, support, and most importantly love, projects such as this would not be possible.

Technical Reviewer and Contributor

Sean Thurston (CCDP, CCNP, MCSE, MCP+I) is a Senior Solution Architect with Seimans Business Services. He provides Network and Data Center design solutions for large-scale deployment. His specialties include implementation of multivendor routing and switching equipment and XoIP (Everything over IP) installations. Sean's background includes positions as a technical analyst for Sprint-Paranet and the Director of a Brick and Mortar advertising dot-com. Sean is also a contributing author to Syngress Publishing's *Building a Cisco Network for Windows 2000* (ISBN: 1-928994-00-8). Sean lives in Renton, WA with his fiancée Kerry. He is currently pursuing his CCIE.

Contents

Answers to Your Frequently Asked Questions

Q: What is *five-nines*?

A: The term *five-nines* refers to an uptime of 99.999 percent. This yields service that is available for all but approximately eight hours per year.

Explore the Four Primary Roles a Server Can Take On in a Cluster

- Primary CallManager Server
- Backup CallManager Server
- Database Publisher Server
- Trivial File Transfer Protocol (TFTP) Server

Understand the Capabilities of Gateway Protocols

Session Initiation Protocol supports five elements of establishing and terminating communications:

- User location
- User capabilities
- User availability
- Call setup
- Call handling

Learn the Guidelines for Multiple Clusters

There are three multicluster designs that may be tailored to fit your design goals:

- Multiple clusters within a campus or Metropolitan Area Network (MAN)
- Multiple clusters over a multisite WAN with distributed call processing
- Multiple clusters over a multisite WAN with centralized call processing

Design a Large H.323 Network

NOTE

As of 12.1(5)XM, the upper level, or directory gate keeper could only service approximately six lower level gatekeepers. As this limit will likely change often, you should check with your local Cisco resource or the Cisco TAC for updated limits.

Understand the Difference between Conferencing and Transcoding

- **Conferencing** is the process of joining multiple callers into a single multiway call. The two types of multiparticipant voice calls supported by the Cisco CallManager are ad-hoc and meet-me.
- **Transcoding** is the process of converting IP packets of voice streams between a low bit-rate (LBR) CODEC to G.711. Transcoding functions can be done by converting G.723 and G.729 CODECs to G.711.

Understand the Advantages and Disadvantages of Using RSVP

Advantages:

- **Admissions Control** RSVP not only provides QoS, but also helps other applications by *not* transmitting when the network is busy.
- **Network Independence/ Flexibility** RSVP is not dependent on a particular networking architecture.
- **Interoperability** RSVP works inside existing protocols and with other QoS mechanisms.
- **Distributed** RSVP is a distributed service and therefore has no central point of failure.
- **Transparency** RSVP can tunnel across an RSVP-unaware network.

Disadvantages:

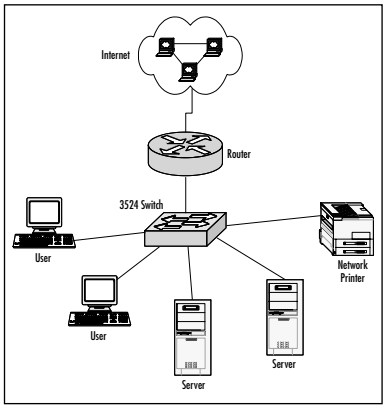
- **Scaling Issues** Multifield classification and statefulness of reservations may consume memory and CPU resources.
- **Route selection and stability** The shortest path may not have available resources, and the active path may go down.
- **Setup time** An application cannot start transmitting until the reservation has been completed.

Designing & Planning…

Dial Plan Preferences:

It is generally considered a good idea to create a dial plan that preferences certain paths routed across the IP network. If this network becomes unavailable, then calls should be routed across the PSTN. As always, the process should be transparent to the user.

A Typical Small Site Traditional Data Network



Use Voice-Capable Gateways

A *voice-capable gateway* is a Cisco router that runs the MGCP IOS firmware that performs processing for voice calls on the local network to local or external destinations. The voice-capable gateways for branch offices are:

- Model 175x for small site gateways, for up to 10 users
- Model 26xx for small sites, for up to 50 users
- Mixed variations of these two devices

Foreword

The business benefits compelling enterprises to assess and deploy IP telephony solutions are many. While many of the cost savings benefits still apply with reduced costs for moves, adds and changes (MACs), infrastructure consolidation, and international toll bypass, perhaps more compelling are the new applications now available. "Click to dial," "follow-me," unified messaging, and customized XML applications running on IP phones can change the way organizations function to gain advantage in the increasingly competitive business landscape.

Many lessons have already been learned from the rapid adoption of Cisco's AVVID IP Telephony product line. While the business justification is compelling, organizations must cost effectively integrate "old world" technologies with the new, engineer packet transport networks for quality of service, simultaneously bolster traditional telephony and IP skills, and continuously exercise proper risk management. The purpose of this book is to help readers overcome the inherent complexities of IP telephony so that its promise can be fully realized.

Those interested in implementing, administering, or gaining certification with AVVID IP Telephony should read this book. A survey of "old world" technologies is presented to contrast to the "new world" of IP telephony and its components. Gateway selection, high availability CallManager clustering, H.323 gatekeeper design, DSP provisioning, dial plans, and QoS are explained. An interesting review of applicable protocols, engineering for complex multiple site deployments, and an overview of IP telephony applications are also included.

The lessons to be learned from this book have been learned by Callisma's highly skilled team of technology and project management professionals who specialize in the design and implementation of complex IP telephony networks. We help our clients leverage competitive advantage through the judicious application of networking technologies via strategic business planning, design, engineering, and implementation services.

—Ralph Troupe, President and CEO, Callisma

Chapter 1

Old World Technologies

Solutions in this chapter:

- Introduction to PBXs
- Looking Inside the PBX
- Interpreting PBX Terminology
- Working with Analog Systems
- Benefiting from Digital Systems
- Providing Video Services

- ☑ Summary
- ☑ Solutions Fast Track
- ☑ Frequently Asked Questions

Introduction

The modern Private Branch Exchange (PBX) system is an old world technology in much the same way an automobile is—there may be better and faster transportation alternatives, but the car's place in daily life, even an older car, is somewhat assured. Cars are used to get people and things from place to place with independence, and they are generally efficient and reliable devices. As with the automobile, the reality is that the PBX in many corporations is not going to depart from the landscape any time soon.

As with the car, telephone systems can be practical or luxurious—ranging from a single line to older key systems to full-function PBX-based conferencing systems with multiple features. The biggest factor in these systems is that a traditional phone is reliable, easy to operate (at least for the end user), and economical. The PBX is a great example of this old-is-good construct, and in many companies this will be true for some time to come. The term *old world* should not be understood as meaning *defunct*.

As such, this chapter will focus on these systems in the context of both their legacy status in telecommunications and their position in newer, more advanced implementations. One of these new models must include Cisco System's vision of the next generation PBX, which is part of the Architecture for Voice, Video, and Integrated Data (AVVID). Please note that Cisco's vision of the next generation PBX *is* the current generation of CallManager and IP phone. The term *next generation PBX* relates to the idea that legacy PBX systems, including Meridian and Sidemen's, will fade into the historical record of telecommunications, following a period of migration.

Under this vision, PBX-attached phones and services will be augmented with Internet Protocol (IP)-based telephony and new integrated services, including video, desktop integration, and data services. Many of these functions and offerings are available today, and a clear roadmap of what can be availed in the next year is easily identified. However, there are still hurdles to overcome, including reliability and simplicity. It is important to remember that while AVVID can provide many features in the network and easily replace the modern PBX in many organizations, the hurdles to this convergence include capital costs, support costs, functionality, training, and reliability.

This chapter will introduce readers unfamiliar with the PBX to these systems and should provide a common reference point for readers already experienced with voice systems. We will also consider the history of telecommunications and

provide a foundation for those entering the AVVID environment of converging voice with data and video information.

Introduction to PBXs

The evolution of phone systems starts with the early experiments of Alexander Graham Bell. In 1875, communication over long distances was handled by the telegraph, a simple device that would transmit electrical pulses across a wire. Though there is some dispute regarding Mr. Bell's status as the inventor of the basic telephone, he and his estate have successfully defended the original patent on the invention. For our purposes, the transmission of sound over electricity is what's significant.

Early telephones were little more than extensions of this original discovery, and party lines and local phone companies were quite common. The party line placed a multipoint drop from the phone company to a number of homes, and the operator would signal an incoming call by altering the ring frequency to a custom signal (for example, one long and three short rings). These systems were prone to the same issues as broadcast media today, especially eavesdropping.

By 1950, a human operator was needed in a more centralized fashion to service the hundreds of phones that were installed. Human operators manually connected calls at the physical layer by using huge switchboards, and call setup times were very long. This system was a solid first-generation effort to link party lines and private lines into a national network. However, many of today's advanced features, including conferencing, alternate billing, and voice mail were inconceivable then.

In the last half of the twentieth century, phone technology made huge strides, including analog switching, digital switching, trunking, and the first versions of the modern PBX. The human operator was replaced with automated switches that processed calls automatically, and corporations were able to provide privately administered services that rivaled the phone company. You may recall that the original public phone systems were virtually always installed and owned by the government or by single corporations—a far cry from the divergent world of today in many countries.

While the PBX remains an entrenched fixture in many organizations, like the mainframe computer, it also gave life to the next generation of successors and augmenters. In the mainframe world this is the personal computer, and in the voice and PBX world this is the IP telephone. Many in the AVVID arena consider the IP telephone a cornerstone, because it is the simplest of devices for the

end user to operate and because it integrates so well with the additional services promised by the technology. Voice over IP (VoIP) is the common term used for these systems. For completeness, and to simplify installation, IP was selected, as it is the most common protocol in the current networking world.

NOTE

In subsequent chapters, you will likely find that many of the problems encountered with VoIP systems, including latency, queuing, and routing, are related to the early decision of using IP as a protocol.

Designing with Legacy Systems in Mind

Before you tackle the converged world of Cisco's AVVID—even if you configure PBX systems daily—it may be a good idea to read this chapter to renew your understanding of what a PBX is and how it works.

NOTE

Please note that this chapter is written from the Cisco AVVID perspective as it relates to PBX systems and telephony, and, as such, some definitions and concepts will differ from the phone company or PBX system origins. These are not errors, but rather, are simplifications of these terms and ideas to a common, related reference point. For example, FXS and FXO in carrier terms can refer to other companies and their respective connections.

However, before we enter the world of the PBX, there is a legacy system that needs introduction. This is the *key system*. A key system is a multiline phone historically found in offices with up to ten users. It is best thought of as those old, clicking phones with the large, lit buttons.

It is possible to find such systems servicing up to 100 users, however, modern economics and the lack of advanced features makes these installations less common, and well-suited for replacement.

As contrasted with the PBX, these systems function by placing a single line on more than one physical phone and, typically, a one-for-one relationship is

maintained between the number of phones and the number of outside lines. As such, unlike the PBX, these systems do not scale to hundreds of users, nor do they save circuit charges.

So, why do we introduce the key system before the PBX? Well, the key system is to the PBX what, presumably, the PBX is to VoIP and AVVID. The services provided by the key system were invaluable to companies of the mid-twentieth century, as calls needed to be routed from one resource to another. In addition, many PBXs today emulate the key system's multiline presence, and this service is available with the current offering of AVVID. As you read about the internal functions of the PBX, consider the legacy of phone and key systems previously described, and consider those services in the VoIP environment.

Designing & Planning…

What Voice Designers Do

The art of voice system design is very different from data installations, although there are similarities. A voice designer is typically confronted with two challenges—the tariff, or cost-per-minute-per-mile, and the redundancy within the network itself. These designs are based on the number of channels needed, and are greatly simplified by the lack of routing protocols and intelligent end-stations.

For example, in a data network installation, the designer will typically draw upon elements of the three-tier model. This model defines a *network core*, which interconnects different *distribution layer* devices, and these, in turn, connect to the *access layer*, which services users. This design is based upon the concept that data packets will take alternate paths between devices based on load, in addition to the premise that the network devices themselves are prone to failure.

Voice designs are different in regards to both hierarchy and redundancy. First, the modern PBX is *internally self-redundant*, which means the physical box itself attempts to provide its own redundancy. Data networking systems have only recently reached this level of redundancy, and, typically, they still experience a short outage as the system changes from the primary to standby engine. In addition, the illusion of redundancy within the box in data networking often requires alterations to the connected devices—Cisco's Hot Standby Router Protocol (HSRP) is a good example of how workstations are tricked into thinking that two

Continued

physically redundant routers are actually one device. The trick is a shared IP address and virtual Media Access Control (MAC) address to make two routers appear as a single router. This, coupled with redundant Supervisor and routing engines, can create the appearance of a redundancy intradevice—however, because the end station has intelligence (unlike the phone), these installations are more complex.

As noted, the end stations in voice networks do not have intelligence, which greatly simplifies the redundancy model. The internally self-redundant PBX, therefore, is not concerned with protocols and other user-side functions to provide redundancy. Within the chassis, a PBX only needs to provide redundant power, redundant processing, and alternate egress paths. Advanced systems may also provide an ability to redirect the physical port to another interface (engine) so the user's phone is also serviced in the event of a hardware failure. This is an uncommon installation, however. Note that all of these redundancies occur intrachassis, and, because of the static nature of the switching paths, no convergence (compared to IP routing) occurs.

By now, you have likely guessed that the hierarchical design considerations in many PBX systems are also very different from routers and switches. For example, it is rare to have a PBX system with three tiers. Most large installations are serviced with two tiers sufficiently. These designs parallel hub-and-spoke data models much more than the three-tier requirements of large data networks. Part of this variance in design is availed by the constant bit-rate of voice and the use of time division multiplexing (TDM). Thus, a designer in a PBX environment need only concern himself or herself with the number of concurrent calls between points. All traffic consumes the same amount of bandwidth (a DS-0 in most cases).

Let's look at that another way. A data designer reviewing the capacity of a link needs two variables—the number of flows and the size of each flow. This is analogous to a freeway where semi trailers use the same road as cars and motorcycles. Clearly the roadway can service more motorcycles than trucks. In contrast, the voice designer needs only one variable in addition to time—the number of flows. All flows are exactly the same—in the highway example, they would all be Volkswagens. Thus, a designer need only consider the number of flows that will occur at the same moment. This may result in a peak of 12 calls at 2:00 P.M.—a figure easily within the capacity of a T-1 circuit including growth and bursts in call volume. The voice designer then adds resiliency and redundancy to the design, in addition to tariffs, or pricing, to develop a network.

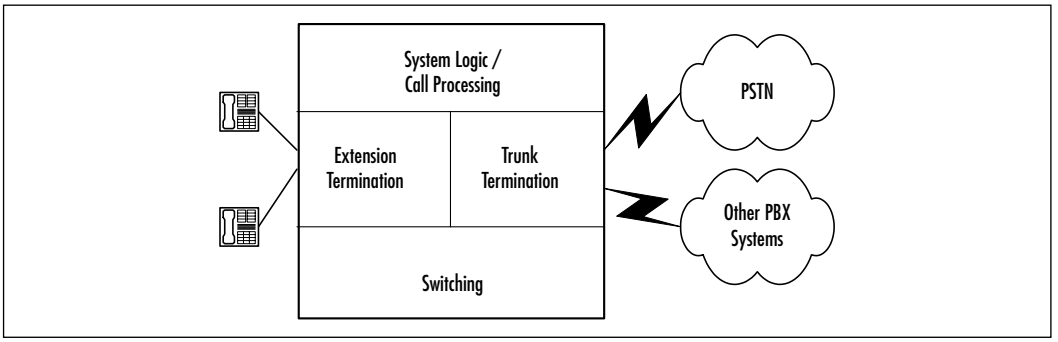
Looking Inside the PBX

A PBX consists of hardware and software designed to emulate the public telephone system within a company, and provide paths into the Public Switched Telephone Network (PSTN). These systems can be categorized into four primary areas, with each area containing one or more functions:

- Extension termination
- Trunk termination
- System logic and call processing
- Switching

These functions are illustrated in Figure 1.1 and described in greater detail in the next sections.

Figure 1.1 The Basic Functions of a PBX



Implementing Extension Termination

Each resource on the private side of the PBX is commonly called an *extension*. These devices have a direct, one-for-one connection to a port on the PBX. These connections are typically digital, however, analog extensions for modems and other services are available, and you will find that the term Foreign Exchange Station (FXS) is commonly used for analog stations such as fax machines and modems attached to the PBX (although this is an erroneous term). In addition, there is a large population of PBXs attached, via analog links, to the extensions, and while the current connections from many vendors are digital, there is nothing wrong with the analog connections apart from the limitations of the transport. Wiring for these connections is voice grade. However, it may include

Category-3 or Category-5, and two- or four-wire (*single pair* or *two pair*) installations are common. The PBX must also provide these extensions with dial tone generation, just as the public phone switch provides this service for non-PBX attached phones. These interfaces also pass the Dual Tone Multi-Frequency (DTMF) tones to the call processing engine that will be described shortly.

Implementing Trunk Termination

While not required, most PBX systems are connected to at least one T-1 circuit for connectivity to either the PSTN or another PBX within the company. A *trunk* is a T-1 or other type of circuit, which can carry multiple channels, or time division multiplexed (TDM) data streams. Recall that these connections can carry up to 24 voice connections depending on their framing and signaling. Please note that trunks can also use the E-1 standard, which allows for 30 user channels.

NOTE

Some trunks are called *tie lines*, which are simply trunks used for connections to another PBX. In some instances these connections are only capable of carrying voice channels; additional functions are provided in others. One example is Siemens' CorNet, which can provide most intra-PBX services between inter-PBX devices.

Call Processing and System Logic

In addition to the user interface found on most PBX systems, there is also logic that controls the flow of calls. The basic process is based on dialing plans, which compare the DTMF tones to the route plans and paths configured on the PBX. These tones represent the numeric values of the buttons, in addition to the asterisk (*) and pound (#) keys. Using the phone number or extension dialed, the PBX routes the call either to the external trunk (the link to the public network), to another PBX within the company (which is carried on an internal trunk), or to another extension within the PBX. This addressing is signaled using the DTMF tones.

The PBX can also make decisions based on its static tables in a dynamic fashion. You're probably thinking this doesn't make sense, but it does. Recall that a PBX route plan specifies the path an outbound call should take. What would

happen if that path failed? Simply, the administrator would specify an alternate path—analogous to a floating static route in Cisco routing. These less-preferred routes could be configured for call overflow (where insufficient capacity exists on the primary link) or trunk failure (where the link must completely fail before taking an alternate path). This decision adds a dynamic to the typically static limitations of the PBX forwarding system.

NOTE

Most PBX phones are digitally connected to the PBX and do not send the actual DTMF tones from the phone to the switch. Traditional analog phones and some PBX phones will send the actual tones to be interpreted by the switch. However, the call routing is still based on the numbers pressed and received, and the non-Signaling System 7 (SS7) signaling is either proprietary or DTMF.

As a designer, you may specify that long-distance calls (indicated with a 9, followed by a ten-digit number, for example) should use a trunk to long-distance provider A, which also provides the lowest cost per minute to the company. The alternate path, configured for overflow calls, might go to long-distance company B, which may also charge more per call. A backup path, using the local exchange carrier, may be configured in the event the first two paths are unusable.

The system logic and call processing functions typically include collections of billing information and other call accounting data that can be used for capacity planning and charge-back services. These functions are independent of the final PBX functional area: switching.

Switching

In order to better understand the diversity of the call routing and circuit switching processes, each is presented as a distinct element in this section. In practice, you will likely find that the two are so inter-related as to be one. In many systems, however, there is a difference.

Switching in the PBX system is the mapping of a channel on one interface to another channel on another interface. For example, this may involve linking a DS-0 to a DS-1 (T-1), or an FXS port to a T-1 trunk on another PBX. The logic that decides which path to be taken is part of the call processing function. Once established, however, the switching of these TDM packets is transparent to the

processor until the call is torn down. This is a significant difference between IP networking and voice traffic, as a routing process typically takes place for each packet—in voice, the call setup only requires processing before the call begins.

It is significant to note that, as with data networking switches, the technology can be blocking or nonblocking and this, coupled with other factors, can greatly impact total capacity. For example, Siemens' blocking architecture can switch up to 5,760 ports, while the nonblocking Intecom can switch up to 60,000 ports.

Establishing Links Outside the PBX

The systems outside the PBX are actually pretty simple once you understand the internal systems. The voice world is made up of *trunks*, which interconnect public or private switches. The basic functionality of these devices is no different for our purposes.

However, there are a few things you should consider when thinking of external PBX resources. These include the wide variety of phone numbers in the international phone network, and the signaling protocol between switches in the public network.

As you may know, calling internationally from your respective country can be either a simple or difficult process. The administration of all the possible numbers is also a daunting task. In either the legacy or AVVID environment, you'll need to work with these external-dialing plans to allow users to connect to other systems.

Consider your home telephone for a moment. In the United States, a call to Israel would require calling 011 (the international escape code), 972 (the international country code for Israel), 3 (the city code, similar to an area code), and the local number, which may be six or seven digits. However, note that in some countries, the city code may appear as 03. A call to Belarus would use a country code of 375, and the city code and number may only contain five digits. A call from another country to the US would require a three-digit area code and a seven-digit number. As a PBX programmer, the system must be capable of handling all the digits provided and routing the call to the correct destination.

Now, with the home phone, the routing of the call is simple—the phone company takes care of it! But, when we enter the PBX, we may have multiple paths to consider. Though this can become very complex, the basics might involve the use of private links between systems (*tie lines*). Consider the United States to Israel example again. It may be cheaper to route calls from Denver to Tel Aviv through the private tie line terminating in Jerusalem rather than the public network, and, although unlikely, it may be cheaper still to route calls for Mozyr,

Belarus, from Denver to Tel Aviv to Mozyr. This dialing plan addresses two factors: call routing and call tariffing.

However, let's presume our call to Mozyr is cheaper using the public network and employing a link between New York and London. How does the network understand our call and establish a path between Denver and Mozyr?

Well, this is the second point of external systems. The switches in the network need to signal each other using a common protocol. In many networks, this protocol is called Signaling System 7 (SS7).

Data network designers are probably used to in-band signaling, where the IP address is part of each packet. No such mechanism exists in voice networks. Rather, the signaling is out-of-band, or independent of the actual data. SS7 is used between the switches to provide this dialog, and, in our call to Mozyr, the Denver phone company switch might use SS7 to signal a path from Denver to Chicago, and another link from Chicago to New York. Once the path is built using SS7, a voice link is established and the call commences. Please note that this does not occur with the PBX private connection to Jerusalem, as this is in-network, and SS7 is typically not used in private switch-to-switch communications.

Configuring & Implementing…

How PBX Installation Differs from Router Installation

Most readers of this book are likely entering into the world of PBX systems from the data world. In fact, many of you may never have installed a PBX or voice system. However, whether you approach data from a voice background or voice from a data background, the reality is that at a high level the two differ less than you might imagine.

It would be inappropriate to enter into the commands and syntax of PBX configuration here—for one, which system would we use as a reference? There are many PBX systems, each with different software versions and hardware options, and each revision of code introduces new commands and syntax. This is not unlike an academic conversation on router configuration—Cisco or Nortel, Multilayer Switch Feature Card (MSFC) or Route Switch Module (RSM) or Route Switch Processor (RSP)-2. In fact, this is the first of the ways in which the systems parallel one another. PBXs and routers both have their own unique features and commands based on the vendor and the version of code.

Continued

In the previous sidebar, "What Voice Designers Do," we discussed the design and deployment considerations of a modern PBX. We also saw the similarities between data systems and voice PBXs. These similarities include redundancy, cost/performance, and design limitations. PBX systems augment these similarities with a few distinct differences, including:

- **Power** Electrical requirements in PBX systems are frequently 48 volt DC. Data network devices are often 120 volt AC.
- **Wiring** It is rare that a PBX system will require Category 5 cabling for connectivity, unlike Ethernet. In addition, it is uncommon to terminate voice grade wiring on patch panels. Rather, voice wiring uses punch-down blocks that hold each bare wire onto a clip. Requirements such as maintaining twists and staying under 100 meters are not part of the typical voice installation.
- **Dial Plan** Unlike IP routers, voice systems rely on static routing tables when forwarding calls. Calls are routed based on a match with the destination number—unlike data networks, the source address is rarely used for call routing. The static route map will define a preferred path, an alternate path, and, sometimes, tertiary paths for each number within the environment.
- **Circuits** In the data world, most circuits are billed at a flat-rate per month. These charges can be distance insensitive (as in the case in Frame Relay), or distance sensitive (common in leased line connections). In voice, it is common to use leased line connections and the associated tariffs, which can allow for significant savings when traffic is carried on alternative paths. These paths may be the connection to the long-distance provider, or may be a private leased line between PBX systems.

Interpreting PBX Terminology

The world of telecommunications and PBX systems includes a vocabulary unique unto itself. You may find that many of the words and acronyms are familiar and common if your background is based in the data world. Nevertheless, there are a number of new terms and concepts that need to be understood before tackling the integration of voice and data systems. In addition, some acronyms have multiple

meanings depending on whether you're discussing voice or data. For example, the acronym *CDP*, to a Cisco router guru, likely means *Cisco Discovery Protocol*. In the voice world this term refers to *Coordinated Dial Plan*.

So, what are the common PBX terms you may encounter? Well, the first is a T-1. A T-1 circuit is capable of carrying up to 24 voice channels (DS-0), depending on provisioning. The total available bandwidth is 1.544 Mbps, although the Integrated Services Digital Network (ISDN) Primary Rate Interface (PRI), which uses T-1 framing, takes one DS-0 for upper layer signaling. The European standard is called E-1. It provides, however, 2.048 Mbps of bandwidth, or 32 channels. An E-1-based PRI, on the other hand, uses two of these channels for signaling and framing, and thus, allows for 30 user-based voice channels. In addition to the T-1 ISDN PRI , the circuit may also be configured as channel associated signaling (CAS) or ear-and-mouth (E&M).

It is warranted to expand on ear-and-mouth technology slightly in this forum, as E&M ports are found on the Cisco hardware platforms and many interconnections will make use of this specification. E&M can also stand for *earth and magneto*, amongst other variations, and is simply another signaling methodology. E&M, like FXO and FXS, is an analog specification, unlike ISDN, which is digital. In addition, FXO is available for PSTN or PBX connections, whereas E&M is for trunk or tie lines between switches—they are network-to-network links. As such, some Cisco installations use the VIC-2E/M interface for connections to voice mail or legacy PBX systems. Please note that this module supports both the two and four wire specifications of E&M for types I, II, III, and V.

These links may also be loop start, in which removing the receiver from the hook closes a circuit and creates a loop, allowing connections. Or they may be ground start, where an earth ground is needed to complete the loop and allow connectivity.

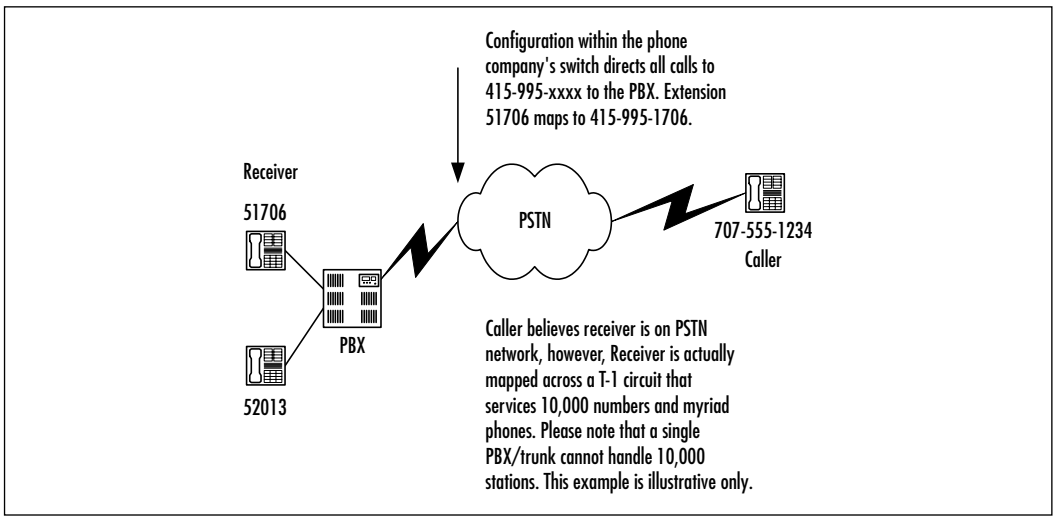
NOTE

It is important to remember that voice services are based on time division multiplexing, or TDM. This is the basis for most connections in the voice world, just as it is for T-1 signaling. A DS-0 is a single voice digital channel of traditional voice bandwidth—8Hz at 8 bits per sample.

The term *central office* is a legacy description of the local telephone company's termination point for all numbers in a given area, and commonly connects to

PBXs via T-1s. Historically these were centrally located and copper was run from each building in the town to the central office. Today, a wide variety of devices are deployed to convert copper local loops into fiber and the central office terminates a small number of fiber pairs that service hundreds of lines. The central office would also provide a Direct Inward Dial (DID), although such connections are typically bi-directional today. In order to directly connect from the public phone system to a PBX, the caller must either be manually routed to the extension or a relationship between the extension and a public number must be established. DID provides the latter service—a block of numbers can be assigned to a trunk line from the telephone provider to the PBX, and the PBX administrator can route those numbers to related extensions. Figure 1.2 illustrates the logical configuration of number 415-555-1234 to extension 51234. Please note that it is quite common to create five-digit extensions in North America that relate to the assigned DID numbers.

Figure 1.2 Logical View of Direct Inward Dialing



To understand the routing in the phone network, one needs to understand Coordinated Dial Plans (CDP). (As we mentioned earlier in this section, if you are entering the world of telephony from the Cisco router, you are no doubt thinking *Cisco Discovery Protocol* for CDP. The acronym CDP stands for both actually, depending on your perspective.) A coordinated dial plan is analogous to addressing in IP routing—the dial plan defines what numbers exist on your network and how callers will reach phones outside your company. (For example, a

coordinated dial plan may require a nine to be dialed before an external number.) The term *call routing* has two meanings, however, that overlap slightly. The first context is the physical act of routing a call through the network. For example, calls to 312 are destined for Chicago, which is a long-distance call requiring a service provider beyond the PBX. The second meaning involves the act of processing that call—there may be three alternate paths to Chicago, and, based on availability, price, and preference, an administrator can route the call along any of those paths.

NOTE

Routing in PBX systems is static, unlike data networks, which typically use dynamic routing.

In the voice world, telecommunications services are billed at various amounts based on the tariff involved. A flat rate structure removes per-minute charges from the billing calculation. Other tariffs can remove distance or other parameters from the calculation.

It is also important to consider the historical import of which end is which in the voice world. For example, you may hear the term *tip-and-ring* in single pair copper connections, which relates to which end supplies the voltage on the wire. In the same manner, there are also *foreign exchanges*, which have slightly different meanings depending on your background.

NOTE

Most voice copper wires are described in pairs—thus, a two-wire connection is a *single pair*, while a four-wire connection is a double pair, or *two pair*. Traditional Category 5 wiring would be *four pair*.

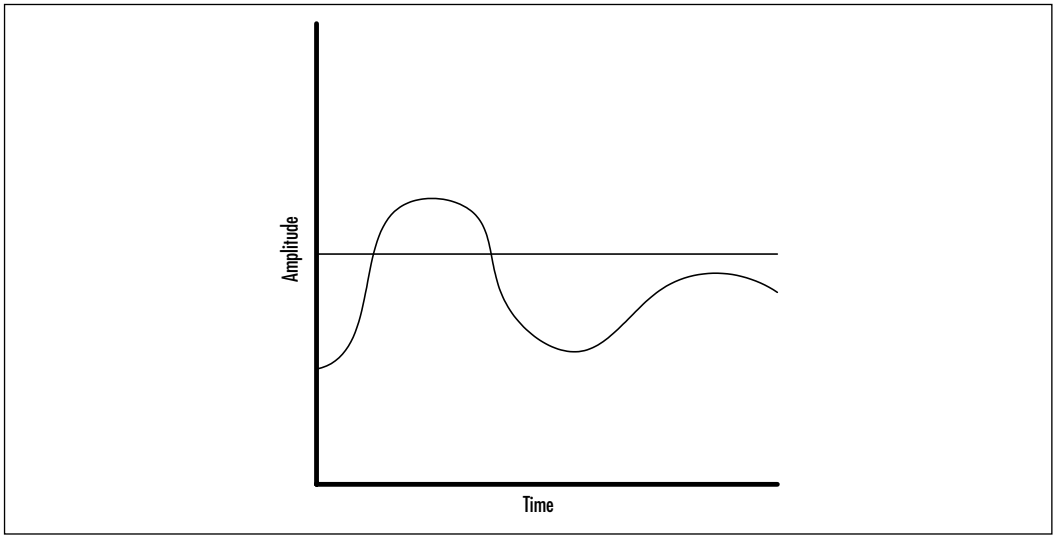
For our purposes, we will describe a Foreign Exchange Station (FXS) as a link between the switch and an extension. This term is sometimes used to describe a connection that services an analog device within the company attached to the PBX, such as a fax machine or modem. If you've worked within the phone company, this term may be defined differently; however, this definition is best in the context of AVVID. In contrast, a Foreign Exchange Office (FXO) link is between

the PBX and the central office. It is a DS-0 and analog, and it is tariffed at a flat-rate. There may be instances when local services are desired, but ISDN or T-1 bandwidths are not needed. FXO connections can be used to service these situations, and can also be used to provide local 911 service in the event all other calls traverse the private network to a main site in another location.

Working with Analog Systems

Analog waves, unlike digital signaling, have a range of values that represent transmitted information. These signals are susceptible to many forms of interference, and, visually represented, they appear as a continuous wave. Figure 1.3 illustrates a common analog waveform.

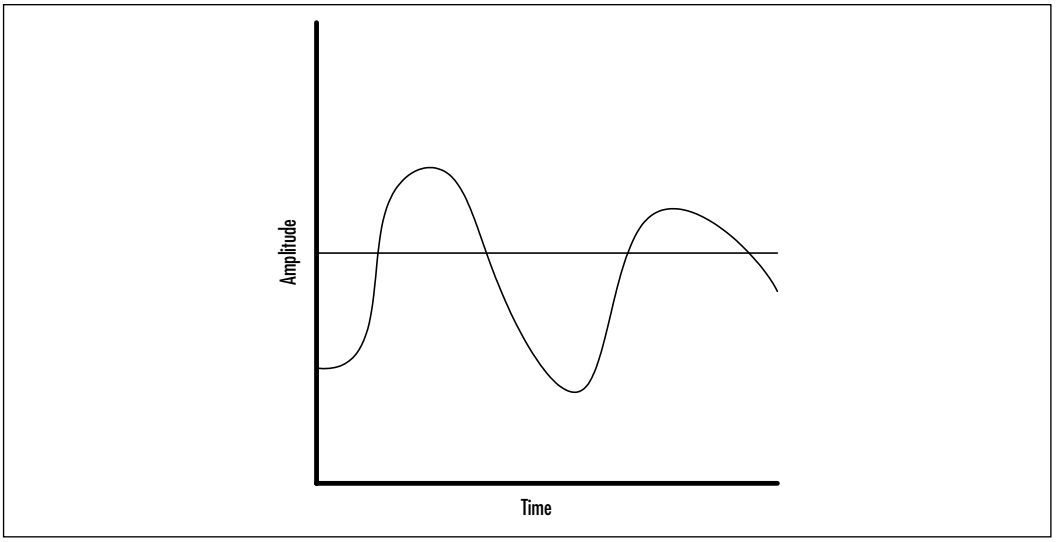
Figure 1.3 The Analog Continuous Waveform



As shown, there is no absolute value within the wave—it varies as the strength of the signal increases or decreases. This introduces one of the primary problems with analog systems, because one must consider the introduction of static and amplification in the waveform. To illustrate this, consider Figure 1.4.

Note that the waveform is now comprised of higher highs and lower lows, and spikes of noise have slightly altered the waveform. The receiver will perceive this as a change in pitch, volume, and tone, and, should this degradation continue through multiple amplifiers and noise-prone circuits, the original waveform may be so disrupted that communications is impossible.

Figure 1.4 The Amplification of Static in an Analog Waveform



The phone system was originally designed to make use of limited frequencies to transmit voice signals. As human speech consumed a very small spectrum, the analog telephone equipment could perform the relatively simple mechanical to electrical conversion necessary to propagate a voice over long distances.

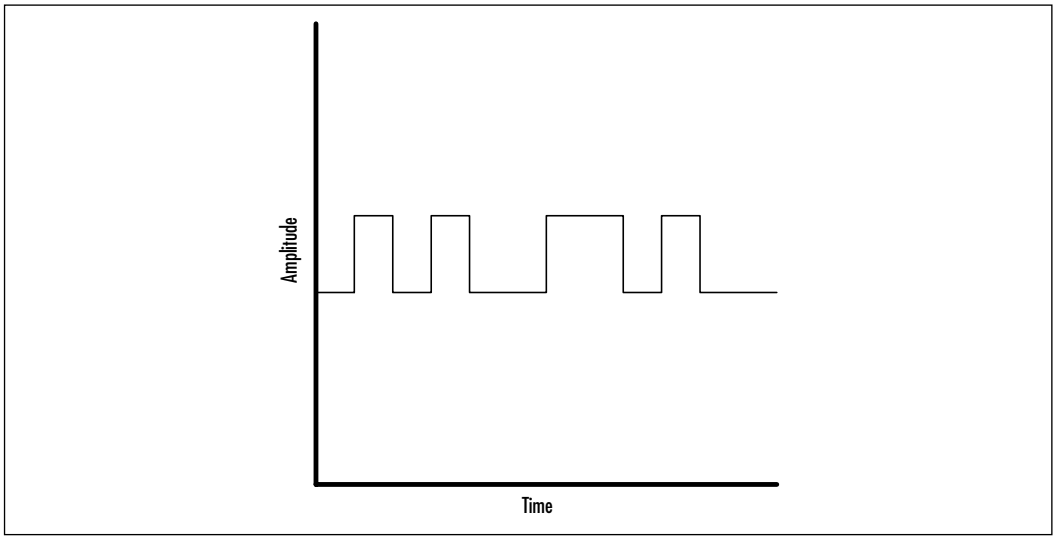
As with record players and compact disc/DVD players, it is likely that both analog systems and their digital counterparts will remain for some time. As such, it is important to consider how analog systems integrate into digital environments such as VoIP or AVVID. Simply put, such installations will require conversions from analog to digital, and, as with old 45s, the quality and performance of the older systems may be limited. Of course, it will also be familiar and, at a political level, you may find reluctance in getting users off their non-VoIP systems.

In the next section we will present digital systems. It needs to be noted here, however, that there is a way to convert from analog to digital—a conversion addressed by a coder-decoder (CODEC). The actual conversion is effectively a sampling of the analog stream and a digital representation of that stream. Of course, the conversion can take the digital data and interpolate an analog waveform. The conversion is not without potential loss, unfortunately, and it is best to limit the number of conversions within a data flow. Recall that FXO, FXS, and E&M are all analog connection methods.

Benefiting from Digital Systems

Digital signals are binary in nature, and are either on or off. These states are very precise, and unlike the continuous waveform that exists in analog systems, the signal can be regenerated with accuracy regardless of noise and interference. This is not to infer that digital signals are impervious to noise and static, but, rather, these problems are easily detected in a digital system and can be compensated for. This is made possible by the absolute values transmitted on the wire. Figure 1.5 illustrates a digital waveform.

Figure 1.5 The Digital Waveform



Digital systems in telephony can take advantage of this binary state and augment communications with additional features that are not available in analog systems, including compression. This allows speech to be sent in fewer bits than in analog format, and, in the migration to AVVID, the data stream can actually be stopped when a party stops speaking. This can greatly increase the volume of connections that can concurrently occur in the network. ISDN PRI is one of the most popular digital connections.

Providing Video Services

It is atypical to include the PBX as part of a video solution; however, some advanced PBX systems do provide video services. These connections can either

be provided over broadband technologies or by way of Ethernet, but it is more common in many systems to use the PBX as a termination point for multiple ISDN Basic Rate Interface (BRI) channels. The BRI can transfer 128 Kbps of user data, and these connections can be combined, or multiplexed, to provide higher levels of bandwidth. Many video conferencing systems work well with 384 Kbps.

In later chapters, we will discuss the technical specifics of the various protocols in use for these connections, including the H.320 specifications, which govern the basic concepts regarding video transmission, including audio and video processing, and are focused on lower-bandwidth media—ISDN and 56 Kbps specifically. This protocol supports point-to-point and multipoint sessions, and provisioning for multicast or multipoint connections is an important consideration in the video environment.

One of the first reactions many users have to compressed video is that it isn't like a television picture. The image is smaller and rougher, and, while it does not have to be so degraded, most vendors haven't forced the additional bandwidth or processing requirements on end users. Adaptation, it is hoped, is to be driven by function, which, in turn, may lead to faster networks and components. This will likely be a slow process, as evidenced by the migration to high definition television (HDTV).

In the United States, the analog video standard is called NTSC, or National Television System Committee. Some in the industry claim that the acronym should stand for Never Twice Same Color, being that, compared to the European and Asian standards, the color information is poorly interpreted from set to set. The NTSC standard specifies a frame rate, or screen refresh rate, of 30 frames-per-second (29.97). Users of these sets are quite accustomed to the grainy picture provided and poor color resolution, and, while HDTV has been available in various forms for years, the FCC and other authorities are already concerned in later 2001 that their 2006 mandate for HDTV conversion will fail. Video conferencing may fail to generate sufficient drivers to make users upgrade their systems, and may exist in degraded form for some time. Or it may also become the next killer-application. This conundrum is a common theme in AVVID, and will be interesting to watch as the old world meets the new.

Audio and video systems require common protocols to define the communications stream, and these standards can be referred to as the H.300s, G.700s, and the T.120s, in homage to the base numbering associated with each standard. This is in addition to the transport protocols of ISDN, Digital Subscriber Line (DSL),

Plain Old Telephone System (POTS), and others. The H, G, and T standards are administered by the International Telecommunications Union (ITU).

The most universal of these video protocols is H.320, which defines a number of parameters including picture size and bandwidth requirements, and will operate within point-to-point and multipoint applications.

It would be unfair to only note H.320 in a discussion of video conferencing protocols, however. H.261, for example, specifies the compression of real-time audio and video data, and defines a screen size of 176 x 144 pixels (Quarter Common Intermediate Format [QCIF]) to 352 x 288 (CIF). Most of these will fit into the bandwidth availed by ISDN. H323 is most often referred to today, and is commonly found in many applications, including the conferencing software provided with Microsoft Windows.

The technicalities of all of these protocols is not important at this point in a discussion of AVVID, and subsequent chapters will elaborate on the standards used by Cisco's CallManager and other resources, such as the IP phones. You will find that many of the protocols used in AVVID telephony are the same as those used in traditional video conferencing, and, because of this, there is integration between the voice applications of the IP phone and the more traditional video conferencing systems such as Microsoft's NetMeeting. For example, one can call a NetMeeting user from a Cisco IP phone.

Summary

Modern telecommunications evolved from the advent of the telephone, and many systems today are related in some way to the first calls. Analog signaling provided continuous waveforms on which data, including voice, could be transmitted. This was replaced with digital signaling, which added compression, error correction, and other services to the network.

As companies grew and their dependence on technology increased, it became necessary to scale the public network more efficiently. This led to the advent of the Private Branch Exchange (PBX). The earliest of these devices provided simple line aggregation, where a company of 1,000 employees could be serviced with 24 telephone lines based on the typical demand for only a small number of concurrent calls. These systems quickly grew in both size and functionality to the point where modern PBX systems can service over a thousand users and provide conferencing and messaging. While these systems have significant acquisition costs, the overall savings provided compared to running single lines to each employee quickly offsets the expense.

Modern PBX is comprised of software and hardware that performs the static routing of calls. These decisions are configured based on links inside and outside the system, which typically include access circuits to the public network, private circuits used between PBX systems in the same company, and lines that service the individual extensions.

Video services grew as an extension of the early analog systems (many television signals are still carried by analog signals into homes). Like the phone system, these systems are evolving, and satellite, cable, and digital subscriber line (DSL)-based services frequently use digital services as well as additional features provided. The signals are converted from analog to digital waveforms via coders/decoders (CODECs).

In computing, the mix of voice, video, and data is commonly called convergence, a reference to the merging of these three data forms into a single transmission medium. That medium is digital and typically based on IP, and will be the subject of the next chapters. Cisco calls this convergence Architecture for Voice, Video, and Integrated Data, or AVVID.

Solutions Fast Track

Introduction to PBXs

- ☑ Private Branch Exchange (PBX) systems provide corporate users with advanced voice services.
- ☑ The modern PBX is a reliable and robust tool on the network.
- ☑ Voice over IP (VoIP) technology is based on the Internet Protocol (IP) because it is the most common protocol in the networking world; however, the choice of this protocol brought with it problems of latency, queuing, and routing.
- ☑ Many PBXs today emulate the legacy key system's multiline presence, and this service is available with the current offering of AVVID.

Looking Inside the PBX

- ☑ The PBX uses trunks and lines to connect to resources.
- ☑ Call switching is distinct from call processing.
- ☑ Each PBX uses a variety of proprietary and standards-based protocols.

Interpreting PBX Terminology

- ☑ Bandwidths are based on analog channels: DS-0 (64 Kbps), DS-1 (T-1, 1.544 Mbps), DS-3 (45 Mbps).
- ☑ Links are called trunks.
- ☑ Some acronyms in the voice world have different meanings in the data world.

Working with Analog Systems

- ☑ Analog signals are continuous waveforms.
- ☑ Analog signals are susceptible to interference and are difficult to correct for errors.
- ☑ Analog signals cannot be compressed without loss.

Benefiting from Digital Systems

- ☑ Digital signals are binary, made up of on or off signals.
- ☑ Digital signals can be compressed, corrected, and manipulated more easily than analog signals.
- ☑ Amplification can occur in digital signals without amplifying background noise and static.

Providing Video Services

- ☑ Video services can demand the most real-time bandwidth in the network.
- ☑ Video data is typically compressed to reduce its load on the network.
- ☑ One-to-many video is a perfect application for IP multicast.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: What is *five-nines*?

A: The term *five-nines* refers to an uptime of 99.999 percent. This yields service that is available for all but approximately eight hours per year.

Q: How are analog signals converted to digital signals?

A: The most common converter is called a digital to analog converter (DAC). This is a specific type of CODEC, or coder/decoder. CODECs are common in the AVVID environment and will be presented in greater detail in future chapters.

Q: How are most AVVID installations deployed today?

A: This varies greatly with the services needed for the company; however, only a handful of corporations are removing the majority of their PBX-based systems in favor of AVVID or alternative solutions. Dow Chemical and Cisco Systems are two of the most publicized, large-scale deployments in mid-2001, but many companies are looking to these systems for pilot or *greenfield* installations. It is becoming common to find portions of a company migrated to AVVID, and many small companies (under 1,000 users) are converting or performing initial installations using AVVID.

Chapter 2

New World Technologies

Solutions in this chapter:

- Introduction to IP Telephony
- IP Telephony Components
- Exploring IP Telephony Applications
- Introduction to Video
- Enhancing Network Infrastructure
- What Does the Future Hold?

- ☑ Summary
- ☑ Solutions Fast Track
- ☑ Frequently Asked Questions

Introduction

The implementation of *new world* telephony technology and video components is something that is viewed with both reverence as well as apprehension. However, that apprehension can be relieved when you consider the potential of implementing such a framework—the benefits of some of the new world applications that Cisco AVVID allows you to use are substantial. When we discuss new world technologies, all we are really discussing are those applications and infrastructures that allow for and build a *converged network*. By converging voice, video, and data together, an organization can enjoy the benefits of simplified administration, toll bypass, and unified messaging, just to name a few. No longer are proprietary systems being tied into a single vendor. We now have this open ecosystem of partners who are allowing you to decide your own fate. If you do not like a specific product, don't use it—use a different vendor, or if you have the ability, develop your own product. Because AVVID is built on many industry standards and open protocols, using different vendors or creating your own solutions has never been easier. Of course, with the advent of these new technologies and solutions we are introduced to an entirely new set of benefits and challenges. The old model of three separate infrastructures and disciplines will no longer serve; we must now change our thinking in certain ways.

This chapter will be covering some basic concepts, defining Internet Protocol (IP) telephony and IP-based video conferencing; we will also be introducing you to the components that make up a converged infrastructure. Although the list seems to grow each day, this converged infrastructure includes CallManager servers, IP telephones, gateways, video-conferencing solutions, and specialized router and switch interfaces so that you can place all of the new world applications such as WebAttendant, IP SoftPhone Unity Messaging solutions, Integrated Voice Response, and AutoAttendant on this converged infrastructure. Our focus will be on introducing the solutions and technologies you will encounter in the design and deployment of your AVVID network, which will be covered in detail in later chapters.

Introduction to IP Telephony

IP telephony is a term used to describe a suite of products and solutions used to transport voice traffic over a data network. Utilizing Internet Protocol (IP) as a transport mechanism, IP telephony allows you to create a converged network in which all communications (voice, video, or data) share the same infrastructure.

There are numerous benefits to this type of infrastructure, including simplified administration, cost savings on telecommunications fees, and unified messaging services.

Simplifying Administration

Almost every mid- to large-sized corporation has a large data infrastructure and along with it, they probably have a large infrastructure built for voice-based traffic. These networks, while both crucial to the organization, share no common thread. Although they may share the same cabling, and even in some cases the same protocols (such as IP), they are still very different types of infrastructures. Two different groups within the corporation administer them, they utilize the equipment of different vendors, both require separate leased lines or plain old telephone service (POTS) lines, and funding for both probably come from different budgets. With the IP telephony solution, these two infrastructures are collapsed into one IP-based network, allowing all communications to share the same administration, ultimately saving time and money for the corporation.

As we discussed earlier, an organization typically has two groups, a *voice group* and a *network group*. Under the old world telephony solutions, these two groups perform very different functions, and in a figurative sense, almost speak different languages. With the IP telephony solution, these groups are collapsed into a single resource pool. Voice and data, while still very different types of traffic, are administered by the same group. Customer service and satisfaction will also benefit from this type of infrastructure. Instead of an end-user having to call the network group for one problem and the voice group for another, the user has a single point of contact for their communication needs.

Utilizing Toll Bypass

One of IP telephony's key features is also one of its most enticing benefits, a feature known as *toll bypass*. Toll bypass allows an organization to utilize its existing data infrastructure to make calls within the organization. Imagine a multinational organization with branch offices spread throughout the world. In the old-world solution, any time one office placed a call to another, the telephone systems of each office would employ the services of telecommunications service providers to place a call within their own organization. If you have ever traveled, you may have experienced the sting of how expensive international calls can be. I placed a call on a business trip from a branch office in Moscow to their headquarters in Cleveland; the call lasted around 40 minutes, and the bill turned out to be $300.00.

So you can imagine how expensive international telecommunications must be for the day-to-day operations of a multinational organization. Now imagine that same scenario using IP telephony, placing that same call from the branch office in Moscow to headquarters in Cleveland, this time utilizing the IP telephony solution. Instead of utilizing the telephone company's services and infrastructure, you would employ the existing leased data lines between the two sites. Now the only price you are incurring is the fixed price you pay each month for the leased line that was already there. As I am sure you can see, IP telephony has the potential to save an organization a great deal of money.

Linking Communications with Unified Messaging

Unified messaging is both one of the goals and benefits of a truly converged network. It links an end-user's voice-mail, e-mail, and fax solutions so they are essentially one entity. With IP telephony, a user could listen to his e-mail, review his voice-mail via software on his PC, review e-mail or listen to voice messages on an IP telephone. Cisco, as well as other vendors, have, and are, developing software applications to utilize unified messaging. We will discuss some of these solutions in the sections to come.

Choosing to Implement IP Telephony

IP telephony sounds great, right? Shouldn't every organization have implemented it by now? Well, first of all, you should keep in mind that voice traffic and regular IP data traffic are two completely different solutions. Regular Transmission Control Protocol/IP (TCP/IP) data traffic is very resilient. It can be forgiving of slow wide area network (WAN) links, lost packets, and the reception of packets out of sequence. In fact, TCP/IP operates in just that way, taking data and segmenting it into several packets and transmitting the data via the best possible path. It is not concerned with the order in which the data is received, or the path it takes to get there, because the end device is responsible for the reassembly and resegmentation of the data. Voice traffic, on the other hand, is not so forgiving, nor as resilient. Even though the voice traffic is being converted to IP packets, it is still voice traffic. IP telephony depends on packets being received in the *same* order in which they were sent; if a packet is lost, then it should remain lost, as retransmitting the packet would only confuse the person on the receiving end of the call. In order to accomplish this, you must incorporate several new features on your routers and switches, such as Queuing and Real-Time Transport Protocol (RTP).

In fact, in order to make IP telephony a reality, your infrastructure is going to need quite a few enhancements. There are several components that must be added to your infrastructure. These components include, but are not limited to, specialized router interfaces, specialized local area network (LAN) switch modules and interfaces, IP telephone handsets, Cisco CallManager servers, and Cisco Unity Mail, as well as other unified messaging solutions. In addition to the required hardware, there are several applications that will also help you to realize the benefits of IP telephony. Applications such as Cisco's WebAttendant, AutoAttendant, and Personal Assistant, as well as third-party software should also be incorporated into your IP telephony solution.

IP Telephony Components

The *components* that must be added to your infrastructure in order to facilitate IP telephony are what really blur the line between the traditional voice infrastructure and your data infrastructure. Here we cross a line into a new realm of devices—but are they voice or are they network? The answer, of course, is that they are both. I think an important point to remember when considering a converged infrastructure is that no matter what we are dealing with, voice, video, or data, it is all communications. This is the information needed for the end-user to effectively carry out his or her business. Perhaps we should begin to consider ourselves communications engineers as opposed to using the traditional network engineer or voice systems administrator titles that have helped to separate the different disciplines for decades. In this section, we will discuss some of these components and their features.

Cisco CallManager

Cisco CallManager provides the IP telephony solution with a software-based call processing platform to fill the role of a traditional PBX. CallManager represents one of the first large-scale enterprise solutions to answer the challenge of IP telephony. As an aside, IP telephony is by no means a new idea. Several companies have introduced VoIP solutions. For example, several Internet Chat programs such as Microsoft NetMeeting, America Online (AOL) Instant Messenger, and Yahoo! Messenger offer the ability to communicate via voice by utilizing the Internet or other network as a medium. While fun to play with, however, it is difficult to imagine an organization utilizing them for an enterprise-wide IP telephony solution, because solutions such as these are essentially entertainment software, and provide for no hierarchy or reliability.

Cisco's CallManager offers a scalable, reliable, and manageable solution for an organization of almost any size and demographic. While it may not be the ultimate choice for IP telephony, it has set a standard of performance for IP telephony call processing, and will probably continue to do so for the foreseeable future. In this section, we will further discuss the CallManager platform, its architecture, hardware, benefits, and limitations.

The CallManager Platform

CallManager is probably the most integral part of Cisco's IP telephony solution. It provides the rest of the IP telephony architecture with a central point for call processing, connection services, signaling, and registration for IP telephone handsets, analog and digital gateways, and legacy telephony devices such as PBX systems. Communication with IP telephony devices is enabled by the use of several IP telephony protocols such as Skinny Station Protocol (SSP), H.323, Media Gateway Control Protocol (MGCP), and Simplified Message Desk Interface (SMDI). These protocols will be discussed in more detail later in the chapter. CallManager offers an open programming interface utilizing the Telephony Application Programming Interface (TAPI) and the Java Telephony Application Programming Interface (JTAPI). By utilizing industry standard protocols, Cisco has opened the door for several other software vendors to further augment the IP telephony product offering. Some of these applications will be discussed later in this chapter as well as in Chapter 7.

Current releases of the CallManager platform allow a single CallManager server to support up to 2500 IP telephone/5000 IP telephony devices per individual server. An IP device can be any of the following:

- IP telephone
- Analog or digital gateway
- IP SoftPhone
- Digital signal processor (DSP)

CallManager has gone through two major revisions. The first revision of the CallManager Platform was the 2.*x* release of the platform. This revision has been discontinued, and CallManager 3.*x* is the current standard, which we will discuss in the sections to follow.

IP Telephony Protocols

In the previous section, we introduced several protocols that CallManager uses to communicate with IP telephony devices. As we discussed in the introduction to this chapter, Cisco is attempting to create an open ecosystem of partners and solutions, with the end goal being to let the organizations decide which product or service best suits them. Supporting several different IP telephony protocols is an important step in this process. It would have been much easier for the Cisco product development team to only support one set of protocols when designing their IP telephony solutions, but by supporting several, they have opened the door to numerous vendors to work within the AVVID framework. Discussed in the next sections, are some of the most common protocols that CallManager can use to communicate. This is by no means a definitive list of all the protocols CallManager will support. As new versions of CallManager become available, the number of supported protocols will also grow. As always, it is a good idea to consult the Cisco Web site for the most up-to-date information regarding this support.

Skinny Station Protocol

Skinny Station Protocol (SSP) is a Cisco communications protocol based on the industry standard Simple Gateway Control Protocol (SGCP) protocol. SSP was first introduced as a method of communication between first generation IP telephone handsets/Gateways (DT-24+/DE-30+) and CallManager servers, and is still widely used today for that same purpose. Products that support SSP include the DT-24 and DE-30 gateways, the Catalyst 6000 8-Port T1/E1 voice service modules, as well as the Catalyst 6000 24 port FXS module. SSP relies on the CallManager server to relay configuration and control information. It is built on TCP/IP and utilizes TCP ports 2000–2002.

H.323

H.323 is an industry-wide open standard for real-time audio, video, and data over packet networks. H.323 is an International Telecommunication Union Telecommunication Standardization Sector (ITU-T) standard and is part of the H.32*x* family of protocols. H.320, transmissions over Integrated Services Digital Network (ISDN), were discussed in Chapter 1. H.323 was built upon this protocol, allowing video and audio transmissions to be supported over packet-based networks such as Ethernet. Cisco's IP telephony architecture can use H.323 to communicate with IP phones, gateways, and, because it is an open protocol, it

can be used to communicate with dissimilar systems such as PBXs and other vendors' equipment. H.323 gateways will be discussed later in this chapter.

Media Gateway Control Protocol

Media Gateway Control Protocol (MGCP) is another Cisco-supported protocol. The CallManager server uses MGCP to communicate with the Cisco VG200 standalone gateway, although several other products in the Cisco product line, including certain products in the Catalyst switching line, will support it soon. MGCP is intended to serve as a faster protocol than H.323 and SSP, utilizing User Datagram Protocol (UDP) as opposed to TCP for transmission. MGCP gateways will be discussed later in this chapter.

Simplified Messaging Desk Interface

Simplified Messaging Desk Interface (SMDI) is the industry standard voice-mail protocol for integrating voice-mail systems with legacy PBX systems and/or other similar devices. CallManager and other unified messaging platforms can use it to integrate with legacy voice-mail systems.

CallManager 3.*x*

CallManager is currently in release version 3.1. The CallManager 3.*x* release introduces several enhancements over the previous 2.*x* version of the software. Version 3.*x* is built on the Microsoft Windows 2000 Operating system, whereas version 2.*x* was built on Windows NT 4.0. Version 3.*x* utilizes a Microsoft SQL server database for data warehousing, while previous versions of CallManager utilized a Microsoft Access database, which severely limited the scalability and reliability of the platform. An important note to make, though, is that CallManager still fails to support other database systems such as Oracle.

CallManager 3.*x* allows up to 2500 IP telephones to be supported by a single CallManager server, up from CallManager 2.*x*'s limit of 200 IP telephones per server. Another enhancement the 3.*x* version of CallManager offers is increased reliability and scalability by use of a feature known as *clustering*. Clustering allows multiple CallManager servers to be interconnected, in order to service more IP telephony devices and to provide redundancy.

Clustering

Clustering will allow you to extend your support for IP devices from 2500 IP telephones on an individual CallManager server, up to a potential 10,000 IP

telephones within a single cluster. Clustering, as its name implies, is the process of combining two or more CallManager servers into a logical unit known as a *group*. A group consists of CallManager servers and their associated devices such as IP telephones, gateways, and logical devices such as SoftPhones, a software-based version of the IP telephone handset. (IP SoftPhones will be discussed further in the IP telephony applications section to follow.) When the group concept is utilized, all the CallManager servers share the same configuration database, so if one CallManager server fails, the others already have the database, thus no manual reconfiguration is required. The idea behind clustering has to do with providing enough servers so that if one of them should fail, the other servers within the cluster can take on the load of the failed server without compromising the level of service to the end systems.

Cisco has outlined four primary roles a server can take on in the cluster:

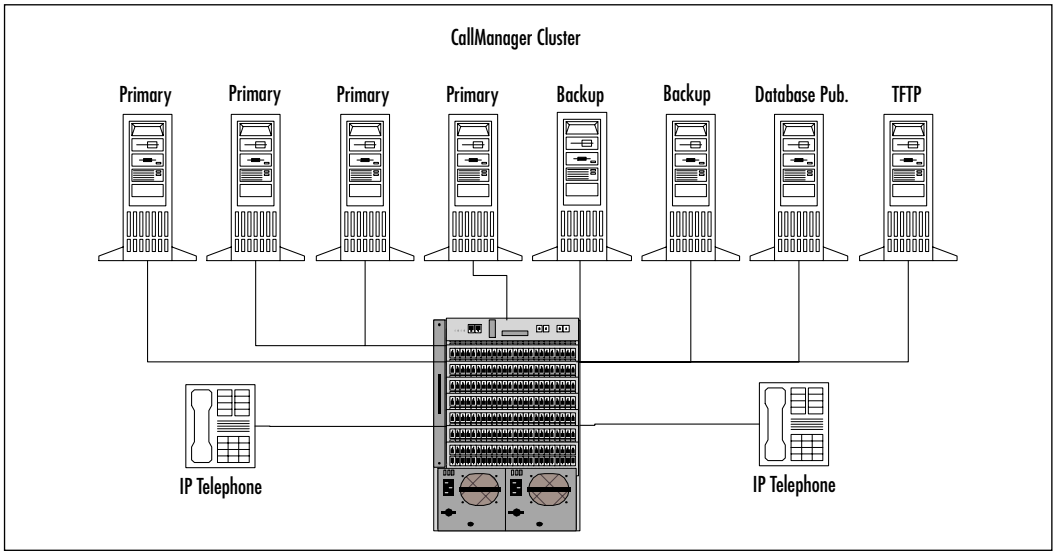
- Primary CallManager server
- Backup CallManager server
- Database publisher server
- Trivial File Transfer Protocol (TFTP) server

The primary and backup CallManager servers are self-explanatory. The database publisher server role is to maintain and distribute the master-configuration database. A second but equally important task is the record warehousing of call detail records (CDRs). A CDR is a record of the IP telephony call. This can be used by other vendors' software for traffic analysis and additional accounting functions. The TFTP server role is used to provide the system image for devices such as IP telephones and gateways.

How you structure your cluster is dependant on how many IP telephony devices will be supported. Cisco has set the following design guidelines for building your CallManager cluster. If you have fewer than 2500 IP telephones, you will need two servers, one primary CallManager server, and one backup CallManager/publisher/TFTP server. For 2500 IP phones, you will need three servers, a primary CallManager server, a backup CallManager server, and a combined Publisher/TFTP server. For 5000 IP phones, you will need four servers, two primary CallManager servers, one backup CallManager server, and a combined Publisher/TFTP server. For the maximum 10,000 IP telephones per cluster, you will need four primary CallManager servers, two backup CallManager servers, one database publisher server, and one TFTP server.

As we discussed in the introduction to this section, there are some limitations you must take into consideration before implementing a cluster. An important item to take into consideration is that a cluster cannot cross a WAN link. All cluster servers must exist on the same LAN. Furthermore, the servers must be interconnected at minimum by a 10 Mbps switched connection. Shared media is not allowed in an AVVID cluster. This is to ensure the proper Quality of Service (QoS) is maintained. Also, as stated earlier, a cluster is limited to 10,000 IP telephones. A maximum of 100 clusters can be interconnected, allowing support for up to 1,000,000 IP telephones within an organization. Figures 2.1 and 2.2 demonstrate clustering and failover protection.

Figure 2.1 CallManager Clustering



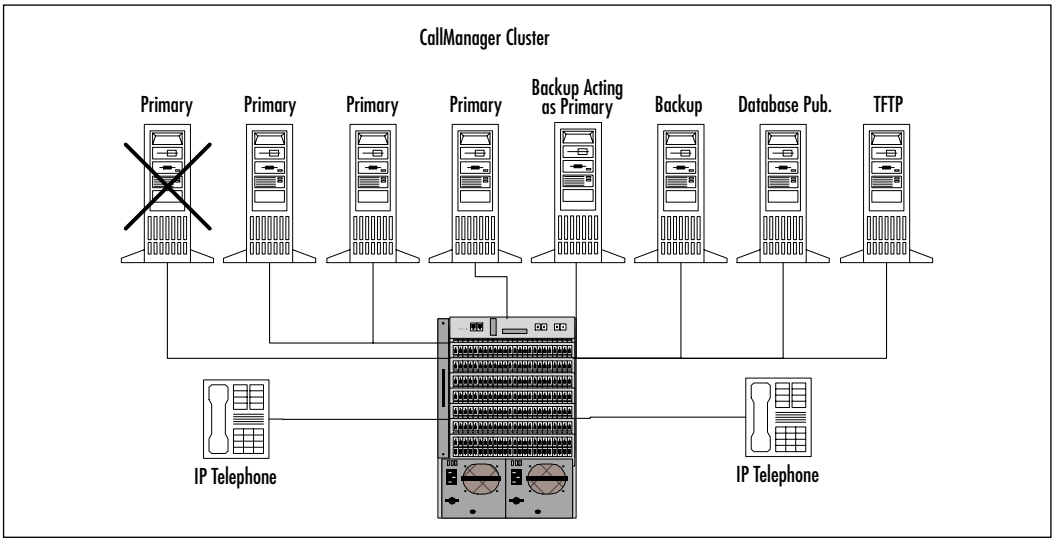
As you can see, clustering gives you a great deal of scalability within your IP telephony network. Making IP telephony a viable solution for organizations ranging from the smallest companies to the largest multinational organizations. Chapter 4 will cover this topic in more depth.

CallManager Hardware

Although CallManager is a software-based application, it must be purchased as part of the Cisco Media Convergence Server (MCS). The MCS servers are Compaq server-class systems. There are several different models of the servers; all essentially perform the same functions—the only real differences are the hardware

features, such as hard drive space, processor speed, and memory capacity. As with all other servers in your network, you should purchase the MCS that best fits your organization's needs. Consult the Cisco Web site (www.cisco.com) for the most up-to-date MCS server information. There is, of course, an exception to the rule of only being able to purchase CallManager preloaded on an MCS server platform. If you have already purchased a Compaq DL320 or DL380 server, meeting specific system requirements as outlined by Cisco, you can purchase a software-only version of the CallManager Software.

Figure 2.2 Failover Protection



WARNING

Because CallManager is a software application, you could probably load it on any server meeting the minimum system requirements for CallManager, although you will probably encounter some amount of difficulty obtaining the software. Should you run into any problems though, you will be on your own. Cisco will not support anything but the approved hardware configurations.

Recently, Cisco announced that the MCS platform will be available on the IBM xSeries of servers as well as Compaq servers. This series of servers will follow the same rules that applied to the Compaq servers, in that the MCS must

be purchased pre-configured. The initial product offering of the MCS platform on the IBM xSeries of servers will be on the xSeries 330 and 340 platforms. I would expect that this group will grow to include other servers in both the IBM and Compaq server lines.

Configuring & Implementing…

What Are the Benefits of CallManager?

Now that we have discussed the specifics dealing with the CallManager, let's discuss the benefits this system will afford your IP telephony solution. As we know, CallManager is a software-based alternative to the traditional PBX system. Traditional PBX systems have the ability to provide an exceptionally high level of service. Cisco CallManager, utilizing clustering technology, has the ability to offer almost these same levels of service and in many cases, CallManager has proven to be an even more reliable alternative to PBX systems. Because it is a distributed system, your call processing functions are protected from a single point of failure, ensuring that your calls can always be made, whereas a traditional PBX system typically offers only a single point of failure.

So, what about the actual features that an administrator and end-user can enjoy? Well, the list of what CallManager offers is quite impressive, although some PBX systems may offer still more services. The list of new services available to CallManager is growing almost daily and is continually being revised and enhanced.

CallManager offers a system administrator the following: SNMP registration, Call Detail Records (CDR), a distributed redundant database, multiple Web-based administration consoles, Dialed Number Identification Service (DNIS), enhanced 911 support, SNMP performance monitoring, and several others.

CallManager offers the end-user the following: call connection and administration, auto-answer of calls, hold and retrieve features, call forwarding, call-park, calling line ID (CLID), Direct Inward Dial (DID), Direct Outward Dial (DOD), distinctive ring service, and several others. This list is growing almost daily as new releases of the software become available.

CallManager is one of the first and arguably the best systems of its kind, offering administrators and end-users an all-in-one IP telephony solution. Scaling from the smallest to largest organizations, it can meet the challenge of almost any environment.

Cisco IP Phones

Cisco IP telephones provide the end-user with an interface into the IP telephony architecture. There have been two generations of IP telephones produced by Cisco: first-generation and second-generation.

Cisco's first-generation IP telephones came with the acquisition of Selsius Technologies. These telephones are now discontinued. There were two models of the first-generation telephones: the 30 VIP/SP+IP telephones and the 12-Series IP telephones, the latter being the most popular. These telephones had a very limited, button-based feature set, while the network interface was a 10 Mbps hub, with an extra interface for a PC or printer. Also, these phones require an external power source, whereas second-generation phones can utilize inline power. Both the 30 VIP/SP+IP telephones and the 12-Series support either G.711 or G.723.1 coder-decoders (CODEC), support Microsoft NetMeeting, H.323 support, and DHCP/Boot P support.

While sharing many similarities with their predecessors, such as support for open standards and the ability to interact with Microsoft NetMeeting, second-generation phones represent a vast improvement over the first-generation phones. Certain second-generation phones interface with the network via a 10/100 Mbps switched connection, also providing an extra port for a PC or other peripheral device, as well as an RS-232 port for additional capabilities. Second-generation phones such as the 7940 and 7960 offer an LCD screen used for a menu-based feature set as opposed to the button-based feature set of their predecessors. The most impressive feature of the second-generation phones is the ability to utilize inline power. Now instead of using an external power supply, these phones, through the use of a specialized inline-power patch panel or specialized modules for the Catalyst switch line, can be powered directly through their category-5 cable. We will discuss inline power options in the infrastructure section later in this chapter.

There are currently four phones in Cisco's second-generation phone offering:

- The 7910/7910+SW phone
- The 7940 phone
- The 7960 phone
- The 7935 phone

Cisco also offers a completely software-based logical IP telephone called the *IP SoftPhone*. The SoftPhone provides an alternative to the hardware-based second

generation IP telephones. It offers a PC-based software application that interfaces directly with the CallManager server to provide IP telephony.

The 7910/7910+SW, 7940, and 7960 are all end-user phones, the only difference really being the features supported, such as menu options, speaker phone, display, and number of lines each phone supports.

The 7960 stands out among its peers as being the only second-generation telephone to offer support for the Station Initiation Protocol (SIP). SIP allows the 7960 to operate without a CallManager on the local LAN. Instead, it communicates directly with the gateway. The 7960 can be expanded further by use of the 7914 expansion module.

The 7914 provides an additional 14 lines to your 7960 telephone, plus two 7914 units can be daisy chained together to provide an additional 28 lines of support. This serves as a great solution for receptionist telephone stations. The 7935 is the speakerphone offering in the second-generation product line. Once again, you should consult the Cisco Web site for the latest product offerings in this line. Table 2.1 discusses the different features of the second-generation IP telephones.

Table 2.1 Second-Generation IP Telephone Features

Features	7910/7910+SW	7940	7960	7935	IP SoftPhone
Network Interface	10 Mbps shared media connection /3-Port 10/100 Switch	3-Port 10/100 Switch	3-Port 10/100 Switch	3-Port 10/100 Switch	Logical interface speed same as the PC it is running on
Number of Lines	One	Two	Six	One	One
Support for the 7914	No	No	Yes	No	No
Speaker Phone	No	Yes	Yes	Only speaker phone	No
Additional RS-232 Port	No	Yes	Yes	No	No
Programmable Keys	No	Yes	Yes	No	No
XML Support	No	Yes	Yes	No	Yes

Continued

Table 2.1 Continued

Features	7910/7910+SW	7940	7960	7935	IP SoftPhone
Support for SIP	No	No	Yes	No	No
Support for Inline Power	Yes	Yes	Yes	Yes	Not needed

Cisco Gateways

Gateways are devices used to connect your IP telephony infrastructure to the Public Switched Telephone Network (PSTN) or to legacy PBX systems. Cisco's product line currently includes over 20 different gateway products, each supporting the various types of gateway protocols. Currently there are three different types of gateways supported by the Cisco IP telephony solution:

- Skinny Gateway Protocol
- H.323
- MGCP

The Skinny Gateway Protocol is based on the industry standard SGCP protocol; however it is only used on the Cisco Gateway product line. In other words, while SGCP is an open standard, the Skinny Gateway Protocol is a proprietary standard used by Cisco only. (This reminds one of the Cisco implementation of High-Level Data Link Control (HDLC)—while HDLC is an industry standard, Cisco has written extensions into it making its implementation inoperable with other vendor's equipment.) Devices that support the Skinny Gateway Protocol include the DT-24+ and DE-30+ gateways, the Catalyst 4000 WS-X4604-GWY module, and the Catalyst 6000 WS-X6608-x1 module.

H.323 is an open industry-wide standard. H.323 gateways are most commonly found in integrated router gateway devices and in communication to Cisco CallManager. Devices that support H.323 include: VG200, the 1750 router, the 3810 router, the 2600 router, the 3600 router, the 7200 router, the 5300 access server, and the Catalyst 4000 WS-X4604-GWY module.

Media Gateway Control Protocol is the most recent of the gateway platforms. MGCP is a Cisco-supported standard and is currently only used in communications between Cisco CallManager and the VG200 standalone gateway, although several members of the Cisco product line will support it in the future. This

group includes the MCS 3810, the 2600 Series routers, the 3600 Series routers, the Catalyst 4000 WS-X4604-GWY module, and the Catalyst 6000 WS-X6608-x1 module. As always, consult the Cisco Web site for information regarding new product support.

Unity Voice-Mail/Unified Messaging Solutions

Unified messaging refers to several products in the Cisco product line that allow end users and administrators to manage all communication from a single point of administration. This product line has undergone several changes within its lifetime, the latest of which came with the Cisco acquisition of the Active Voice Corporation in 2000. With this acquisition, Cisco is offering the Unity product suite as its unified messaging solution. The previous unified messaging solution was a product line known as uOne, which has been discontinued.

The Unity product line is a powerful collection of tools that allows a user to retrieve e-mail, voice-mail, and faxes all from one location—a truly converged solution. Like the rest of Cisco's IP telephony product offering, the Unity product suite is continually being revised. We will discuss some of the features available as of this writing, but bear in mind that new features are most likely to appear in the near future.

Unity integrates with Microsoft Exchange server and the Outlook Mail client to provide a centralized application where a user can retrieve e-mail, voice mail, and faxes. This solution allows users to send, receive, and manage voice messages directly from the Outlook client. Unity also gives users the ability to send and receive faxes directly from the user's Outlook mail client. A user can either fax directly or send e-mail that will be received in the form of a fax. Prior to the acquisition by Cisco, the Unity product line offered an integrated fax solution known as Active Fax. This product is no longer in production. In order to utilize a fax solution with the Unity product suite, a third-party fax server such as that from RightFax or Omtool must also be purchased. Consult the Cisco Web site for the most current listing of approved fax server software.

A personal Web assistant is also included with the Unity suite, allowing users to manage their voice-messaging options directly from a Web page. Users have the ability to change passwords, greetings, mailbox options, and so on, taking the burden off the system administrators and providing users the ability to make changes to their systems as they see fit.

This type of solution provides users with a great deal of flexibility and mobility. A company executive called away on urgent business at the last minute

could use her laptop to check her voice-mail and fax, and she could use the personal Web assistant to update her greeting, letting everyone know she is not in the office.

Exploring IP Telephony Applications

Legacy PBX and similar systems have set a very high benchmark for reliability, scalability, and service. In order for IP telephony to become a viable solution and to either eliminate and/or compete with these systems, the same levels of service and available features must be achieved. Cisco and other vendors, such as Interactive Intelligence, Latitude Communications, and Intelligent Telemanagement Solutions, are developing a number of applications to meet this challenge. The sections that follow discuss a number of these applications and their features and benefits.

Introducing Cisco's IP Telephony Applications

Cisco and other vendors have developed software solutions to further enhance their IP telephony solutions. Along with the opportunities they are fostering, of course, come new and difficult challenges. When we think about Cisco Systems, the first thing that comes to mind is probably not the role of a software vendor, but the world leader in networking hardware. IP telephony applications allow Cisco to augment their IP telephony hardware with features and services to make IP telephony an even more viable solution for Cisco's customers. In the following sections, we'll describe Cisco's WebAttendant, IP SoftPhone, Internet Communications Software (ICS), Interactive Voice Response (IVR), and AutoAttendant services.

Cisco Web Attendant

Cisco WebAttendant is designed to replace traditional manual attendant consoles. It is a Web-based Graphical User Interface (GUI) that allows the user to receive and dispatch calls from any IP phone within the network. WebAttendant works on a client server architecture that allows the IP phone in use to interface directly with the CallManager to direct calls and to monitor the status of lines, much like a traditional receptionist console.

Another added benefit of WebAttendant is the ability it provides system administrators to perform system maintenance from that same easy-to-use Web-based GUI as opposed to the interface of the legacy PBX systems. WebAttendant offers many of the same features offered by traditional PBX systems such as *hunt groups* and multiple attendant consoles.

WebAttendant is included as part of the basic package when purchasing CallManager 3.*x*. It has the ability to scale to meet the size of almost any IP telephony infrastructure. A single WebAttendant console can monitor up to 26 calls at a time. A single CallManager cluster utilizing WebAttendant can support up 32 hunt groups with 16 members per hunt group. Also, a cluster can support up to 96 WebAttendant consoles. That means up to 512 (96 consoles x 26 calls) calls at one time.

When designing your infrastructure to include WebAttendant, make sure to take into consideration all the design limitations discussed in the previous paragraph, such as number of hunt groups (32), number of members within those hunt groups (16), as well as the maximum number of simultaneous conversations possible (512). Your design should never reach the limitations of the WebAttendant system—if you are approaching these design limits you should consider utilizing multiple CallManager clusters.

NOTE

One of Cisco's partners, Arc Solutions (www.arcsolutions.com), is also producing an attendant console software package. While similar to WebAttendant, it offers a more feature-rich and scalable platform.

Cisco IP SoftPhone

Cisco IP SoftPhone is a client-based application that integrates seamlessly with Cisco CallManager, and is designed to allow users to utilize IP telephony from any network-attached PC. All the client requires is a microphone and speaker, and they now have a fully functional IP telephone handset. A GUI on the user's PC provides a dial-pad and other functions present on a standard IP telephony handset. This application provides a great solution for traveling users who need the benefits and features of IP telephony, but are unable to take a regular IP telephony handset with them. An important note to make regarding IP SoftPhone is that it consumes 20 device units on a CallManager server, as opposed to the one used by a standard IP telephone handset. Another note to make regarding IP SoftPhone is that it must be installed with Microsoft NetMeeting—SoftPhone will not work without it. If you are planning to deploy IP SoftPhone on more than a limited basis, ensure that your infrastructure is equipped adequately for the load it will face.

Internet Communications Software

Internet Communications Software (ICS) is a suite of five tools designed for service and application providers to further grasp the benefits of IP telephony. These components are:

- Automatic Call Distribution (ACD)
- Cisco IP Contact Center (IPCC)
- Intelligent Contact Management (ICM)
- Customer Interaction Suite
- Network Applications Manager (NAM)

Automatic Call Distribution

Automatic Call Distribution (ACD) is a tool used to reroute calls to different customers serviced via the same central office. ACD is provided as part of the Network Applications Manager (NAM), which will be discussed later in this section.

Cisco IP Contact Center

Cisco IP Contact Center (IPCC) is an IP telephony solution that allows call centers using IP telephony to receive regular POTS calls as well as IP telephony calls. IPCC can provide the following features: intelligent call routing, computer telephony integration, integration with legacy ACD, and integration with legacy as well as IP-IVR.

Intelligent Contact Management

Intelligent Contact Management (ICM) is due to be released in the first part of 2002. It is a software solution used for direction and relay of customer contact information between resources. This system will utilize a set of user-defined roles in order to route voice, World Wide Web (WWW), and e-mail correspondence to the appropriate system or resource.

Customer Interaction Suite

Customer Interaction Suite is an IP telephony solution that allows corporations and service providers the ability to interact with their customers on the Internet or network in a real-time manner. There are four components to the Customer

Interaction Suite: Cisco Media Manager, Cisco Media Blender, Cisco E-Mail Manager, and Cisco Collaboration Server:

- **Cisco Media Manager** works with **Cisco Collaboration Server** to direct a customer to the resource that will best serve their needs or requests.
- **Cisco Media Blender** does just what its name implies; it blends the different types of media into one format. Voice, text, and WWW traffic can all be combined into one medium, offering a significant cost savings over the traditional model of separate dissimilar systems used to manage customer data and communications.
- **Cisco E-Mail Manager** is used to direct received e-mail to the appropriate party or resource. This allows an organization to cut down on lag time from the moment when e-mail is sent to the organization and when the organization is able to respond to the e-mail.

Network Applications Manager

Network Applications Manager (NAM) is the software solution that gives organizations the ability to utilize all the other ICS components we have just discussed. It provides a hierarchical structure providing a range of services from very simple to very complex. NAM has a long list of benefits and features, including ACD, CTI, IVR, customer relationship management (CRM), Web collaboration, e-mail response management, and call management. Consult Cisco's Web site for the most current information on NAM as well as other IP telephony applications.

Interactive Voice Response

Interactive voice response (IVR) is a voice application designed to handle calls on systems serving as voice-gateways. This system is available in two packages, either as a router equipped with VoIP interfaces and feature sets, or as a server-based Java solution running on Windows NT/2000 servers. The server-based solution is the newest and most feature-rich offering for IVR within the industry. This system offers a Web-enabled GUI management interface, with an open programming customizable model. IVR is used to provide information in the form of voice in response to a user-initiated string of information such as spoken word, key-tones, or telephone line signaling. A very practical application of this solution would be a prepaid calling card system. In such a system, a user would enter a calling-card number and personal identification number (PIN). IVR could be

used to allow/disallow the call, report to the user the number of minutes left on the card, and so on. For more information on IVR and its uses/capabilities, refer to the Cisco Web site.

AutoAttendant

AutoAttendant is a Cisco application that works with IVR and CallManager software to provide call routing services. It allows CallManager to receive calls on specific extensions and then forward that call based on caller input. This type of system could be found in organizations that utilize menu-based systems offering caller options such as dialing a user's extension and/or dial-by-name systems.

Third-Party IP Telephony Applications

As we discussed earlier, Cisco is a networking hardware designer and manufacturer, not a software company. Its primary focus is, and should be, the hardware aspect of IP telephony. Because Cisco's AVVID architecture is built on open standards, it has opened the door for numerous vendors to either write new software to become interoperable with Cisco's solutions and/or to make their existing software interoperable. It seems to have worked. Although IP telephony is a still a relatively new technology, companies are already seeing its potential and have started to develop applications designed to work alongside Cisco's IP telephony architecture. This can only continue to make IP telephony a more accepted alternative to the traditional systems. This section will introduce three vendors who have designed software to work with the IP telephony solution: Interactive Intelligence, Latitude, and ISI. Chapter 7 will discuss these as well as other applications, and how to choose the appropriate applications for your needs.

Interactive Intelligence's Solutions

Interactive Intelligence (www.inin.com) has an Original Equipment Manufacturer (OEM) agreement with Cisco, in which Interactive Intelligence's Interaction Center platform will be included on the Cisco ICS 7750 platform. The Interaction Center platform of software provides a single platform to integrate voice, fax, e-mail, Internet text-chats, WWW requests, and VoIP calls. Interaction Center was designed to run on top of Windows 2000 and includes several different software components. As with most similar software solutions, Interaction Center runs on a client/server architecture, with software installed on both a central processing server and on each Interaction Center client. Interactive Intelligence has also created three specialized versions of the Interaction Center

platform: Customer Interaction Center (CIC), Enterprise Interaction Center (EIC), and Service Interaction Center (SIC).

Latitude Communication's Solutions

Latitude Communications (www.latitude.com) has developed a specialized e-conferencing platform that integrates with Cisco CallManager. Their product is known as MeetingPlace IP. MeetingPlace IP is a client/server based video-conferencing application for mid- to large-sized enterprise environments. MeetingPlace IP offers real-time collaboration applications used for video-conferencing, training, and project management.

Intelligent Telemanagement Solutions

Intelligent Telemanagement Solutions (ISI, at www.isi-info.com) is the first company to introduce an IP telephony accounting application. This is a function that legacy PBX systems and similar devices have been performing for several years, which IP telephony is still far behind on. The ISI system allows administrators to further utilize the benefits of IP telephony toll-bypass, by allowing an administrator to analyze traffic patterns and optimize their infrastructure based on their findings. The ISI system works with the CallManager system, utilizing the CDR for each IP telephony call.

Introduction to Video

Traditional old world video transmissions typically consist of one to several ISDN basic rate interface (BRI) lines connecting proprietary video-conferencing end-stations. These ISDN lines typically operate in a point-to-point infrastructure utilizing the H.320 specification. Usually the bandwidth used is anywhere from 128 Kbps to 384 Kbps, and is kept completely separate from the existing data and voice infrastructures, which results in a large under-utilization of available resources. Although some advanced PBX systems can terminate the BRI lines for the video conferencing systems, the BRI lines and voice lines are kept completely separate from one another. As the technology has improved over the last several years, this type of system has gained a great deal of popularity and it is not uncommon to find some form of this system in most mid- to large-sized organizations.

New world IP-based video conferencing systems allow you to utilize your existing data networking infrastructure as opposed to working with a separate infrastructure, resulting in much better utilization of your network resources.

IP-based video conferencing, on the other hand, utilizes the H.323 specification (discussed earlier in this chapter), allowing you to utilize video conferencing over a variety of mediums including shared and switched media such as Ethernet, leased lines, and nonbroadcast multiaccess networks such as Frame Relay and Asynchronous Transfer Mode (ATM).

As part of the AVVID line of solutions, Cisco offers several solutions to enable video-conferencing to meet the varying needs of organizations of a range of sizes. In the following section we will discuss IP-based video conferencing in greater detail, as well as some of the components used for IP-based video conferencing.

Understanding Video Components

As we discussed in the Introduction to IP Telephony section, VoIP is very intolerant to delay and dropped packets. The statement is even more true when we discuss IP-based video-conferencing or video over IP. Just imagine if you were watching a video conference that was not received in real-time, perhaps a sales presentation or some type of training, and the information was received out of sequence—you could be looking at a chart that you heard about five minutes ago. IP-based video transmissions as well as IP telephony are very similar in nature. Voice, or in this case, video data is encapsulated into IP packets and transported to the end destinations. In the following sections we will discuss some of the components needed to facilitate IP-based video conferencing, such as gateways, gatekeepers, multi-point control units (MCU), video terminal adapters (VTA), and endpoints. We will also briefly discuss the IP/TV product line and the services it provides. As with the rest of the AVVID product offerings, it is highly recommended you consult Cisco's Web site for the most up-to-date information on the products and solutions that we will discuss in this section.

Gateways

Gateways are used to provide you with IP-based video conferencing network access outside of your network. They provide protocol translation, such as H.323 to H.320, and translation to ISDN from other network mediums. For a gateway solution, Cisco offers the IPVC product family. Currently Cisco is offering the IPVC 3520, 3525, and 3540 platforms. These are modular platforms offering LAN, ISDN BRI, ISDN Primary Rate Interface (PRI), and V.35 connection options. As this line is growing rapidly, I would expect that more product offerings are just around the corner for the IPVC 35xx product family.

Gatekeepers

A Gatekeeper is a device used to permit or deny requests for video-conferences; they are an integral part of the IP-based video conferencing solution. It is responsible for deciding if enough resources are available for the video conference to occur, and if the device requesting the conference can gain access to the requested resources. Currently, there are two solutions in the Cisco product line that offer Gatekeeper functionality, the IPVC 3510, and the Multimedia Convergence Manager (MCM) IOS feature set available for the 2500, 2600, 3600, 7200, and MC3810 platforms.

Multi-Point Control Units

A Multi-Point Control Unit (MCU) serves as a center for video-conferencing communications and infrastructure. It serves as a single point of control governing the establishment, joining, and termination of video transmissions. An MCU is needed whenever three or more participants need access to the same real-time video conference. A single MCU can also control several different video conferences simultaneously.

Currently, Cisco offers the IPVC 3510 MCU and IPVC 3540 Multipoint Conference Unit (MCU module) platforms to fill this role. The 3510 can support up to 15 participants in either a single conference or multiple conferences, whereas the 3540 can support up to 100 users in either a single conference or multiple conferences.

Video Terminal Adapter

The Video Terminal Adapter's (VTA) role in video conferencing is to provide an interface to legacy video-conferencing systems. This is accomplished by providing a protocol translation between the legacy H.320 specification for video-conferencing over ISDN and the IP telephony H.323 protocol. As we discussed in the Introduction to Video-Conferencing, many mid- to large-sized organizations have already invested in video-conferencing technology. Utilizing VTAs, these organizations can protect their investment in legacy equipment while still enjoying the benefits of the new IP-based video-conferencing solutions. Currently, Cisco offers the IPVC 3530 platform for VTA functionality.

Endpoint Devices

Endpoints are the end-user devices that subscribe to and receive services from video-conferencing. Cisco does not manufacture an endpoint series of devices;

however, their video-conferencing solutions do support many of the industry standard endpoint devices that support the H.323 specification. This list currently includes Microsoft, PictureTel, Polycom, Sony, TANDBERG, VCON, VTEL, and Zydacron. While systems vary from manufacturer to manufacturer, you will typically find the same components, usually a video camera, video screen, and audio components. Usually manufacturers differentiate themselves by offering better resolution or screen refresh time, while the core functionality for each unit is generally the same. The list of supported vendors is growing almost daily. Consult Cisco's Web site as well as your chosen vendor's Web site to ensure endpoint compatibility.

Cisco IP/TV

The Cisco IP/TV product line is a hardware and software solution designed to provide real-time one-way video broadcasting services to desktop computers. There are two components to this product offering, the IP/TV series of servers and the IP/TV desktop software. The system differs from typical video-conferencing systems in that it utilizes multicast traffic to allow several subscribers to view the same presentation from a single source. This system is often utilized for employee training or company-wide conferences in which only a few parties speak.

The IP/TV server family includes five servers. All are preloaded Windows 2000 servers: the 3411, 3422, 3423, 3431, and 3415. The 3411 serves as a management and broadcast control server. It is responsible for scheduling, server access, balancing of network resources, and control of video services. The 3422 and 3423 servers are responsible for the actual capture, storing, and transmission of live or archived video broadcasts. These servers receive their direction and control from the 3411 server. The 3431 server is an archive server. It is responsible for the storing and cataloging of prerecorded video-transmissions such as training material. This material can then, at any time, be retransmitted by the 3411 and 3422/3423 servers. The 3415 server is the video starter system. It provides an all-in-one IP/TV solution for small organizations that are just getting started with IP/TV. It offers control, broadcast, and storage facilities. While offering an all-in-one solution, it is not intended to replace, nor can it offer the same functionality of the 3411, 3422/3423, and 3431 servers. Rather, it is intended to be a stepping-stone into the larger environment.

The client side of the IP/TV system is a software application known as the *IP/TV viewer*. This software communicates directly with the 3411 control server to attain information regarding available broadcasts and program listings. When

the appropriate program is selected, the IP/TV viewer allows the user to view the broadcast.

Enhancing Network Infrastructure

As we noted for both IP telephony and IP-based video-conferencing, creating an AVVID-enabled network requires a great deal of new equipment. Depending on the needs of the network in question you will most likely be adding devices such as CallManager servers, IP telephones, WebAttendant consoles, video gateways, gatekeepers, MCUs, VTAs, and endpoints. All of these devices (and more) are very necessary to make AVVID a reality for your network. In fact, they require even more additions. Several enhancements also need to be made to your existing infrastructure, such as specialized router interfaces and specialized switch cards. As we discussed before, we must blur the line between our voice and data networks. Here at Layer 2 (the access layer) of the Open Systems Interconnection (OSI) model and Layer 3 (the network layer), where data has always reigned as the proverbial king, we must now make our infrastructure voice *and* data friendly—a shared kingdom of sorts. Previously, we've focused on the upper layers; now we'll discuss the Layer 2 and Layer 3 devices that will make our new type of network a reality.

Using Routers for a Converged Network

As we all know, a router is a Layer 3 device, the primary purpose of which is path determination and packet switching based on IP or other Layer 3 addresses. When we introduce a converged network, routers are going to have to be one of the first places we begin to make enhancements. Cisco has developed several routers that allow a network to make the change to a converged network. Several new types of interfaces have emerged, utilizing the modular chassis capabilities of Cisco's newer routers. Now both voice and video interfaces are available for these routers. In the sections that follow, we will discuss these interfaces and the routers that support them.

Analog Voice Interfaces

Cisco routers utilize *analog* voice interfaces to interface either directly with telephone handsets, or to connect to legacy PBX or the PSTN. Because analog technology is considered a much older and more stable technology, these interfaces are standardized. There are currently three types of analog interfaces supported by

Cisco routers: Foreign Exchange Station (FXS), Foreign Exchange Office (FXO), and ear-and-mouth, sometimes known as earth and magneto (E&M). Let's discuss these interfaces in more detail.

Foreign Exchange Station

Foreign Exchange Station (FXS) ports use a standard RJ-11 telephone jack to connect to telephone handsets, modems, or fax machines. This is the common type of interface found in homes. Cisco routers would most likely use this interface for phone-to-phone connectivity.

Foreign Exchange Office

Foreign Exchange Office (FXO) ports also utilize a standard RJ-11 telephone jack. FXS ports are commonly used by businesses to connect their legacy PBX systems to the service provider's telephone network. Cisco routers can use an FXO port to connect to a legacy PBX device or to directly connect to the PSTN.

Ear-and-Mouth

Ear-and-mouth (E&M) offers a more advanced solution than either the FXO or FXS ports, as well as several features that the other two do not, such as trunking and either analog or digital transmission. E&M utilizes an RJ-48 port as opposed to the RJ-11 used by the others. Cisco routers would most likely use an E&M port for connection to PBX or PSTN, as well as a connection requiring trunking.

Digital Voice Interfaces

Digital voice interfaces are provided to Cisco routers by use of digital voice trunking cards and Digital Voice Processor (DVP) voice compression modules (VCMs). Digital voice trunking cards interface most commonly with ISDN BRI and PRI lines. By utilizing the individual channels on each line, it allows for a single line to support two voice lines using BRI and up to 23 lines using PRI in the U.S., and up to 30 in Europe. Digital voice processor VCMs allow a router to take a voice conversation and compress it down to as small as 5.3 Kbps, depending on the method utilized, as opposed to a 56 Kbps channel. This allows for a much greater utilization of available bandwidth.

MC 3810 Router

The Multi-Service Access Concentrator 3810 (MC 3810) represents the first router of its type, offering the full capabilities of a router as well as Voice over Frame Relay, ATM, and leased lines. It was designed to be an all-inclusive solution for branch-office deployments. A major disadvantage of the MC 3810 is that it is expensive and the network modules used in it are not interchangeable with any other platforms. This is no longer a very popular platform due to the VoIP capabilities of routers such as the 2600 and 3600.

1750 Multi-Service Series Routers

The 1700 Series of routers provide a small office solution for organizations. As a member of the 1700 Series family, the 1750 multiservice router series also offers an IP telephony solution, two analog voice channels, a DSP, and three network interface module slots for additional voice/data support. The 1750 can share the same WAN and voice interface cards as the 2600 Series. This router would most likely serve the small and home office market, due to its small capacity and limited features—it would not be adequate in the larger branch office role.

2600 Series Routers

The 2600 Series of Cisco's routers has become one of the most popular connectivity solutions for branch office connectivity. It offers a modular design, sharing network modules with the 1600, 1700, and 3600 Series of routers, providing two WAN interface card (WIC) connections as well as one network module slot. The 2600 router supports 10 Mbps and 100 Mbps Ethernet interfaces as well as token ring. The 2600 supports VoIP applications and support for up to 48 digital voice lines (60 in Europe). Voice interface cards (VICs) allow the 2600 Series to support analog voice interfaces. By using two VIC cards in the WIC card slots, the 2600 can support up to four analog lines.

3600 Series Routers

The 3600 Series bears many resemblances to the 2600 Series, but the 3600 Series is quite a bit more powerful, offering a great deal more scalability and processing functions. There are three classes in the 3600 Series: the 3620, 3640, and 3660. The 3620 provides two expansion module slots, the 3640 offers four, and the 3660 offers six. Whereas the 2600 supports the use of WICs, the 3600 Series supports the use of carrier cards that provide service for WAN, LAN, and voice interfaces; these cards are interchangeable with the 2600 and 1750 Series routers. LAN support for the 3600 supports 10 and 100 Mbps Ethernet, as well as token ring.

7200 Series Routers

The 7200 Series is Cisco's first-level enterprise router. It offers a four- or six-slot configuration with interfaces including ATM, Synchronous Optical Network Technologies (SONET), ISDN BRI, ISDN PRI, T1, E1, T3, and E3. It also supports AVVID applications through use of the multiservice interchange (MIX) functionality. The MIX allows the 7200 to support digital voice as well as gateway functionality through the use of two different trunk interfaces, the high-capacity and medium-capacity T1/E1 trunk interface cards. The primary difference between the two cards is that the high-capacity card includes an on-board DSP card for compression. The 7200 Series can support up to 120 voice calls depending on the module configuration used. This router also supports analog voice applications through the use of voice interface cards (VICs).

Cisco Switches

Cisco's Catalyst switch line is a highly-advanced line of switching solutions that scale to meet various business needs, from small organizations to multinational corporations. Catalyst switches operate at Layer 2, but Layer 3 switching is also possible with a Route Switch Module (RSM). All of the switches in the Catalyst line support AVVID networks, including IP phones, but specific switches within the product line are designed specifically to meet the needs of IP telephony, specifically inline power, which we will discuss in the next section, and gateway functionality. We will discuss three lines that meet this challenge: the 3500, 4000, and 6000 Series.

3500 Series Switches

The 3500 Series is a scalable, entry-level solution for small- to mid-sized networks. It is a wholly Cisco-developed switch, which runs a router-like IOS. The 3500 Series of switches are fixed configuration switches, all offering 10/100 Ethernet ports and Gigabit Interface Converter (GBIC) ports. The difference comes in the number of ports offered and the forwarding rate at which the switch can process packets. Currently the 3524XL-PWR is the only switch in the 3500 Series that supports inline power, although other models within the 35xx product line will, in the future, most likely offer inline power as well. The 3524XL-PWR offers 24 10/100 Ethernet ports, 2 GBIC ports, and a packet-forwarding rate of 6.5 million packets per second.

4000 Series Switches

The 4000 Series is a step up from the 3500, offering a modular configuration in four different switches: the 4003, 4006, 4840G, and 4908G. The 4000 Series also offers supervisor engine functionality, similar to that of the 5500 Series. Within the 4000 Series, the 4006 is currently the only switch to offer inline power; by use of the Catalyst 4000 inline power 10/100BaseT switching module or the use of an auxiliary power shelf, the 4006 supports up to 240 10/100 ports. The 4003 is also an Ethernet switch, very similar to the 4006, but unfortunately the 4003 cannot offer inline power functionality. The 4840G and 4908G are both gigabit Ethernet switches. The 4000 Series also offers voice-gateway functionality through the use of the Series 4000 WS-X4604-GWY module, which provides support for both H.323 and SSP (in the future it will support MGCP).

6000 Series Switches

The 6000 Series is a highly scalable, enterprise class series of switches. The 6000 Series offers a completely modular design, utilizing supervisor modules, with the capability for redundant supervisor modules, if necessary. There are five switches in the 6000 Series family: the 6006, 6009, 6506, 6509, and 6513. The 6006 and 6506 offer six slots while the 6009 and 6509 offer nine slots. The 6513 is the largest form-factor in the product line, offering 13 slots. The 6000 Series provides inline power directly through the use of specialized 48-port switching blades. The 6006 and 6506 can support up to 240 10/100 ports, while the 6009 and 6509 can support up to 384 10/100 ports, and the 6513 can support up to 576 10/100 ports. Gateway functionality is provided via the WS-X6608-x1 module. This module supports SSP and will in the future support MGCP. The 6000 Series also offers an eight-port voice T1/E1 and services module to provide connectivity to legacy PSTN or PBX systems, as well as a 24-port FXS module for analog telephone connectivity.

Exploring Inline Power Options

During our discussion earlier in this chapter concerning IP telephones, we discussed how second-generation phones were superior to their first-generation counterparts because they offered support for *inline power*. First-generation telephones were limited in that they required an external power source in order to function. Inline power allows second-generation phones to avoid this pitfall.

There are two ways in which inline power can be offered to second-generation telephones, either by way of a power patch panel, or through the use of inline

power modules installed directly in the switch. Let's discuss these different power options as well as their advantages and disadvantages.

Inline Power Modules

Inline power is currently available for three switches in the Catalyst product line: the 3524XL-PWR, the 4000 Series, and the 6000 Series. The 3524XL-PWR is a fixed-configuration 24-port switch. It provides out-of-the-box inline power support. An important note to make is that the 3524XL-PWR switch offers no inline power redundancy.

The 4000 Series provides inline power through use of the Catalyst 4000 Inline power 10/100BaseT Switching module and the Power Entry Module (PEM). Redundancy is provided to the 4006 by use of the WS-P4603 auxiliary DC power shelf. This allows for an N+1 protection scheme protecting against a single power supply failure. An important note to make is that the 4003 cannot interface with the power entry module and therefore cannot utilize inline power directly from the switch. In order for the 4003 to provide inline power, you must use the Catalyst inline power patch panel.

The 6000 Series provides inline power by use of a 48-port switching blade. Inline power is provided to the switch via 2500-watt power supply. Two power supplies can be used for redundancy. Inline power modules offer a great solution in environments where space is at a premium.

Because all of the functions are collapsed into one piece of equipment, administration is simplified. On the down side, this solution may require a forklift upgrade, as inline power modules are only available for the 3500, 4000, and 6000 Series, which could introduce a great deal of added expense. Even though power-redundancy is available for these switches, you are still relying on a single point of failure should the entire switch fail.

Power Patch Panel

The Catalyst inline power patch panel offers an alternative to the forklift upgrade that might be necessary in order to accommodate inline power. This solution allows you to utilize your existing switching infrastructure, such as 2900 and 5000 Series Catalyst switches, by providing inline power external to the switch. The Catalyst inline power patch panel offers 96 ports, for support of up to 48 stations per panel, one port for the IP telephone and one port for the switch. The major advantage to this solution is that it helps to protect your existing investment in switches and helps to keep your options open to future product offerings. The

major disadvantage is that you now have an additional piece of equipment in order to administer.

Power Cube

Power cubes are an external power supply, used as a sole means of power for the first-generation telephone offerings. Power cubes can be used by second-generation telephones as a sole means of power, or more commonly, can be used as a backup power supply to the inline power patch panel and the inline power modules. The advantage to this solution is that, when used with inline power, it provides a redundant power supply for your IP telephone. When used solely as a means of power, its advantage is that you can deploy IP telephones and not have to replace your switches or install inline power patch panels for power. The major disadvantage is that you must provide a power outlet for each cube, and it adds to the mass of cables around a user's desk.

Different Queuing for Video/Voice

Queuing is an important design and performance issue that must also be examined when discussing IP telephony. Queuing has traditionally been a Layer 3 function for WAN connections, but when discussing a converged network, specifically that dealing with voice or video traffic, attention must also be given to the LAN. Layer 2 traffic can be classified by type of service using the 802.1Q protocol. It is recommended that when using this protocol you separate voice and video traffic from regular data traffic and place this traffic in a higher-priority queue. 802.1Q specifies seven classes of service (COS), 0 being lowest priority and 7 being of the highest priority. It is recommended that COS 4–7 be used for voice and video, and that 0–3 be used for normal data operations. An important note to make regarding Layer 2 queuing is that once the packet encounters a router, the Layer 2 information is lost—in other words, 802.1Q is only a LAN solution. For traffic crossing WAN links, Layer 3 queuing must be incorporated.

Designing & Planning…

The Cisco Three-Layer Model

Many people reading this book are probably already familiar with much of the Cisco product line and solutions framework. As a review for those individuals and an introduction to others, now would be a good time to discuss the Cisco three-layer hierarchical model and how it relates to an IP telephony and IP-based video-conferencing network. This model provides an outline for network design, by designating three layers (the *access*, *distribution*, and *core* layers) within a network, each with specific roles and responsibilities that it should carry out.

- **Access Layer** The access layer is where end-nodes and workgroups gain access to the network. Equipment found at this layer usually includes switches such as the Catalyst 2900 or 3500 Series and routers such as the 2600 and 3600. This is the layer in which virtual LANs (VLANs) are implemented, and it is primarily concerned with providing access; policy control and filter are carried out at the higher layers. Regarding IP telephony and IP-based video-conferencing, this is the layer at which we would find IP telephones, video-conferencing end-points, inline power enabled switches, inline power patch panels, and voice-enabled routers. Depending on the size and layout of the network, CallManager servers may also be found at this layer.
- **Distribution Layer** The distribution layer's primary concern is the aggregation of access level connection and policy control. The distribution layer connects the access layers together and provides security by implementing access-lists and encryption. Devices found at this layer are usually routers and high-end switches. Concerning IP telephony and video-conferencing, you would most likely find voice-enabled routers, gateways, gatekeepers, CallManager Servers, and IP/VC or IP/TV servers at this layer.
- **Core Layer** The core layer is the center of the network. This layer's primary objective is the high-speed transfer of data. The core layer should not be concerned with policy control and it should not be allowing single nodes access. Technologies used at this layer typically include ATM,

Continued

SONET, or Gigabit Ethernet. Products found at this layer are typically high-end routers such as the 7500, or Layer 3 switching devices like the 12000 Gigabit Switch Router (GSR). Concerning IP telephony, we would probably not see much here, as it should be implemented at the lower levels.

What Does the Future Hold?

When we discuss the future of AVVID, it is useful to examine where we are and how we got here. Over the past few years, this technology has grown by leaps and bounds. Advances have been made in every area, with new offerings coming out almost every day.

An area that will definitely be enhanced further will be the CallManager platform. In version 2.*x* organizations were limited to 200 IP telephones per server. When 3.0 debuted, this number went up to 2500 with support for 10,000 IP phones within a cluster. With the latest release, 3.1, we can now support up to 1,000,000 IP telephones. I would expect that this number will continue to rise with the later revisions of the software. As this technology gains further acceptance, we can also probably expect to see several enhancements to the CallManager feature offering, such as a voice-recording system. I would also expect to see support specifically for call centers. IOS-based versions of the CallManager platform are already available; although these currently provide a very limited feature set, future revisions will most likely contain more features and open IP telephony solutions up for branch offices.

Aside from CallManager, I would also expect that we will see further software developments and new packages from both Cisco and other vendors. A good example would be the Intelligent Contact Management (ICM) suite, due to be released in the first part of 2002. It is a software solution used for the direction and relay of customer contact information between resources. This system will utilize a set of user-defined roles in order to route voice, WWW, and e-mail correspondence to the appropriate system or resource. This and other systems of its kind will further enhance and augment the IP telephony and AVVID solutions.

Another area we will likely see growth is with the offering of *pizza box solutions*. These products derive their colorful name from their small form-factor, which is about the size of a pizza box. These are integrated all-in-one access solutions that provide capabilities such as routing, switching, and voice-gateway services. A good example of this type of solution would be the IAD1101 integrated

access device. This solution offers two T1 ports, a v.35 port, a 10bastT Ethernet port and an RS-232 port. Eight analog ports are also available via an expansion slot. This type of solution offers several capabilities of the larger scale platforms at a fraction of the cost.

At the client end, I would expect to see a third generation of IP telephones offering wider support for the SIP protocol as well as other features, and maybe eventually integrating with a video endpoint device, although Cisco has not made any indication of this.

Without our standard-issue crystal ball, we cannot be certain as to what the future holds, but if past performance is any indication, I would expect we will continue to see rapid development of exciting products and services. Perhaps IP telephony will one day replace our existing telephone infrastructure. Only time will tell.

Summary

Internet Protocol (IP) telephony and IP-based video conferencing solutions present many opportunities to your organization, and at the same time introduce an entirely new set of challenges to overcome. IP telephony benefits your organization by providing simplified administration, toll bypass, and a unified messaging platform. All of these benefits have the potential to save your organization a great deal both administratively and monetarily when implemented correctly.

Several new IP telephony-specific enhancements need to be made to your infrastructure in order to make IP telephony a reality. The CallManager system replaces the traditional Private Branch Exchange (PBX) system for call processing; currently in release 3.1, it has the potential to support up to 2500 IP telephones per server and 10,000 per cluster. IP telephone handsets provide the user interface to the IP telephony infrastructure. Currently in their second-generation, Cisco offers four IP telephone handsets: the 7910/7910+SW, 7940, 7960, and 7935. *Gateways* provide the interface to the public switched telephone network (PSTN). Cisco products currently support three protocols: the Skinny Station Protocol (SSP), H.323, and Media Gateway Control Protocol (MGCP). The Cisco Unity product suite offers a unified messaging solution, integrating voice, video and fax communication into one medium.

IP telephony applications are used to enhance the IP telephony product offering. Cisco has developed several IP telephony applications to work within the AVVID product offering. WebAttendant is a software-based attendant console used to replace the traditional PBX attendant console. It provides call monitoring and management functions and can be used to monitor up to 26 conversations concurrently. IP SoftPhone is a software-based version of an IP telephone, providing a viable alternative for traveling users or others who do not have access to an IP telephone. Internet Communications Software (ICS) is a grouping of five tools: Automatic Call Distribution (ACD), Cisco IP Contact Center (IPCC), Intelligent Contact Management (ICM), the Customer Interaction Suite, and Network Applications Manager (NAM), for service and application providers. Interactive voice response (IVR) and AutoAttendant are used for menu-based telephone systems. Other vendors have also developed applications to work with Cisco's IP telephony solutions, which includes Interactive Intelligence, Latitude, and Intelligent Telemanagement Solutions (ISI). Interactive Intelligence offers the Interaction Center platform, installed on the ICS 7750. Latitude offers the MeetingPlace IP software for video-conferencing, and ISI offers an accounting and billing application that utilizes the call detail record (CDR) of each call.

Video over IP (VoIP) is made possible through the use of several devices including gateways, gatekeepers, multi-point control units (MCU), video terminal adapters (VTA), and endpoints. Gateways provide access to the outside world from your internal network. Gatekeepers are used to permit or deny requests for video conferences. MCUs serve as a center for video-conferencing communications and infrastructure. The VTA's role in video conferencing is to provide an interface to legacy video-conferencing systems. Endpoints are the end-user devices that subscribe to and receive services from video-conferencing. Cisco offers the IPVC family of products to meet the video-conferencing needs of organizations. Cisco also offers the IP/TV family of products for one-way video broadcasting.

Additional infrastructure, including switches and routers, will also need to be adapted in order to meet the needs of AVVID. Cisco's voice-ready routers include the 1750 Series, 2600 Series, 3600 Series, and 7200 Series. All of Cisco's switches support IP telephony, but the 3524XL-PWR, 4000 Series, and 6000 Series support inline power. Inline power can come in the form of inline power modules for the Catalyst switches or external power-patch panels.

Queuing and class of service (COS) measures must also be taken on your LAN/WAN in order to ensure real-time delivery of voice and video traffic. 802.1Q provides a Layer 2 queuing solution for your LAN environment.

Solutions Fast Track

Introduction to IP Telephony

- ☑ Simplified administration is achieved by converging three separate networks into one, allowing one resource pool to administer the entire network.
- ☑ Toll bypass allows organizations to avoid costly telecommunications expenses by utilizing the data infrastructure.
- ☑ Unified messaging combines voice-mail, e-mail, and faxes into one easy-to-use interface.

IP Telephony Components

- ☑ CallManager provides the IP telephony network with a software-based PBX system.

- ☑ IP telephones provide the user interface to the IP telephony network.
- ☑ Gateways provide the interface between the IP telephony network and the public switched telephone network (PSTN) or a legacy PBX device.

Exploring IP Telephony Applications

- ☑ WebAttendant replaces the traditional PBX attendant console.
- ☑ IP SoftPhone provides a software-based IP telephone handset.
- ☑ Third-party applications include software from Interactive Intelligence, Latitude, and ISI.

Introduction to Video

- ☑ Traditional video-conferencing utilizes ISDN lines in a point-to-point infrastructure.
- ☑ IP-based video-conferencing utilizes the H.323 specification allowing for video-conferencing over a variety of mediums.
- ☑ IP-based video-conferencing is much more efficient than traditional video-conferencing because the existing data infrastructure is utilized opposed to a separate infrastructure.
- ☑ Gateways provide access to the outside world from your internal network.
- ☑ Gatekeepers are used to permit or deny requests for video conferences.
- ☑ Multi-point control units (MCU) serve as a center for video-conferencing communications and infrastructure.

Enhancing Network Infrastructure

- ☑ Routers provide gateway services and voice aggregation for IP telephony by use of analog ports, FXO, FXS, E&M as well as digital trunking cards.
- ☑ Routers that support IP telephony include the 1751, 2600 Series, 3600 Series, and 7200 Series.

- ☑ Switches that support inline power modules include the 3524XL-PWR, 6000 Series, and 4000 Series.
- ☑ Inline power is also provided by using the Catalyst inline power patch panel.

What Does the Future Hold ?

- ☑ Future revisions on CallManager include a call center solution.
- ☑ Pizza box and integrated access devices will provide all-in-one functionality for branch offices.
- ☑ IOS-based versions of CallManager will further develop.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: What benefits can I expect from using IP telephony?

A: IP telephony has several benefits, including simplified administration, toll bypass, and unified messaging.

Q: Can I install CallManager on any server platform?

A: No, CallManager can only be purchased as part of the MCS hardware platform, which is either a Compaq or IBM server platform. A software-only version may be purchased if you have existing Compaq servers that meet Cisco's requirements. Consult the Cisco Web site for these requirements.

Q: Wouldn't it just be easier to purchase the IP SoftPhone application instead of deploying IP phones?

A: Although that solution may provide some up-front cost relief, IP SoftPhone is intended to augment the IP telephones, not replace them. IP SoftPhone does

not have nearly as many features as an IP telephone, and it takes the resources of 20 IP telephone handsets when used. The short answer would be no!

Q: What router should I use for my VoIP solution?

A: The answer to this question is dependant on several factors, such as features needed, number of users/nodes supported, and the potential growth of the location. The best answer would be to pick the router that meets your current needs, but one that will allow for substantial growth and expansion.

Q: I already have Catalyst 1900 and 5500 switches deployed. Do I have to get rid of them and purchase switches that support inline power?

A: No, by use of the Catalyst inline power patch panel, you can protect your existing investment while still providing inline power to your IP telephones.

Chapter 3

AVVID Gateway Selection

Solutions in this chapter:

- Introduction to AVVID Gateways
- Understanding the Capabilities of Gateway Protocols
- Choosing a Voice Gateway Solution
- Choosing a Video Gateway Solution
- Multimedia Conference Manager Services

- ☑ Summary
- ☑ Solutions Fast Track
- ☑ Frequently Asked Questions

Introduction

Gateways are part of the network platform's components, which is a sublayer of the network infrastructure layer of AVVID (Architecture for Voice, Video, and Data). Moving to an AVVID architecture requires integrating with an existing PBX (Private Branch eXchange) and PSTN (Public Switched Telephone Network) infrastructure. Whether there are existing analog or digital voice circuits in place, gateways need to be implemented to merge the legacy architecture to an IP-based voice, video, and data network. In addition, voice gateways provide connectivity from the new AVVID infrastructure to legacy voice mail systems. The gateway to be selected will depend on the type of circuits, the protocol type, and capacity requirements.

Gateway selection is a topic often overlooked, and mistakenly so. When designing an AVVID network, careful consideration should be given to the type of gateway you need to implement within an AVVID network; certain platforms support particular gateway protocols, with each of the protocols providing specific functions that may or may not be required within your network design.

In this chapter, we will be discussing the different types of gateways supported within an AVVID network, looking at the technology from a voice perspective, as well as explaining the IP/VC 3500 range of products that in simple terms will provide H.320 to H.323 conversion.

Introduction to AVVID Gateways

A *gateway*, by definition, is a device that converts one media or protocol to another. In the AVVID or Voice over IP (VoIP) environment, a gateway is responsible for connecting an IP telephone network to the PSTN or PBX and key systems. For example, the gateway may connect an H.323 network to an SIP-based network, PSTN, or ISDN. It also performs translations between different transmission formats and communication procedures, and is responsible for setting up and clearing calls on both sides. Communication between terminals and gateways is done through the H.245 and Q.931 protocols.

Types of gateways range from specialized entry-level standalone devices to enterprise-level integrated router and switch gateways. Based on the device or the implementation, the gateways communicate with Cisco CallManager or other network devices over various gateway protocols. Your own infrastructure and VoIP requirements will help determine what gateway is right for you, but required common features include: DTMF relay, CallManager redundancy, and

supplementary services. Supplementary services allow users to perform call hold, transfer, and conferencing.

AVVID gateways themselves are in the form of analog and digital versions running different protocols, each of which we'll cover in the coming sections.

Understanding the Capabilities of Gateway Protocols

The three voice gateway protocols supported in Cisco's AVVID architecture are Skinny Station Protocol (SSP), H.323, and MGCP. Skinny Station Protocol allows a *Skinny client* to use TCP/IP to transmit and receive calls and RTP/UDP/IP packets for audio. An example of a Skinny client is an IP phone or gateway. The Skinny clients communicate with a Cisco CallManager over TCP on ports 2000–2002. SSP was developed by Cisco (formerly Selsius) as a low-bandwidth gateway protocol.

H.323 is the most supported gateway protocol from Cisco, and is an ITU-T (Telecommunication Standardization Sector of the International Telecommunications Union) standard for packet-based audio, video, and conferencing. It is the umbrella standard for the conferencing standard (made up of others such as H.245, H.225, and Q.931), and is the only gateway that provides full routing capabilities. It transmits and receives media streams via RTP with Real-Time Control Protocol (RTCP), carried over UDP, thereby providing status and control information. Signaling, such as Registration, Admission, and Status (RAS), H.245, and Q.931 is transported over TCP. Q.931 signaling, meanwhile, is for call setup and termination. Capabilities, however, are exchanged by utilizing H.245, which is for call control, and establishes multimedia communication or call services between the H.323 clients. Some new features to be supported in H.323v2 are H.235, H.450.x, Fast Connect, alternative gatekeepers, Q.931 forwarding, and integration of T.120. For its part, H.235 allows security and authentication, such as registration passwords, while H.450.x offers supplementary services.

The MGCP protocol functions in an architecture where the call control intelligence is removed from the gateway. Level3, Bellcore, Cisco, and Nortel developed MGCP (described in Request for Comments [RFC] 2705), which is a master/slave protocol, where the gateway is the slave servicing commands from the master, which is the call agent. In the Cisco AVVID environment, the CallManager functions as the call agent. Two benefits MGCP provides over H.323 are centralized dial plans and dynamic dial plan updates versus statically configuring each H.323 gateway. (The MGCP protocol communicates over UDP port 2427 and TCP port 2428.)

The skinny gateways are the DT-24+, DE-30+, and Catalyst 4000/6000 modules, which provide CallManager access to digital gateways. An example of an H.323 gateway is a Cisco IOS router like the 2600 and 3600. The VG-200 is an MGCP gateway with future support for the 2600, 3600, 3810, and Catalyst modules.

Another protocol being implemented in Cisco gateways is the Session Initiation Protocol. SIP (described in RFC 2543) is an application-layer control protocol that can establish, modify, and terminate multimedia sessions or calls. These sessions include IP conferences, telephone calls, and multimedia distribution. A Cisco VoIP solution for SIP consists of a SIP agent, 7960 IP Phone, SIP gateway, and a SIP proxy server.

SIP supports five elements of establishing and terminating communications:

- User location
- User capabilities
- User availability
- Call setup
- Call handling

Currently, the VoIP world is dominated by H.323; the emergence of SIP and the increasing number of applications supporting this new technology means the interoperability of SIP with existing H.323 networks. SIP was first available in IOS version 12.1(1)T. SIP was enhanced in IOS version 12.2(2)XA, supporting the Cisco 2600 and 3600 Series router-voice gateways, and the Cisco AS5300, AS5350, and AS5400 access server-voice gateways now allow both SIP and H.323 to fully coexist, as well as allowing interoperability between the two protocols.

NOTE

An example of new software utilizing the functionality of SIP is the application Windows Messenger, which is part of Windows XP. Windows Messenger is real-time communications software that provides end-to-end IP telephony. SIP is a multipart communication protocol, but the first version of Windows Messenger will only support a two-way conversation.

Video gateways are used to convert form H.320 devices to H.323 devices. The Cisco IP/VC products allow companies to utilize their legacy H.320 ISDN-

based videoconferencing to integrate with newer IP-based H.323 videoconferencing devices. Cisco IP/VC gateways support H.261 and H.263 video coders/decoders (CODECs); H.261 is used as multiple channels of 64 Kbps, while H.263 is a higher quality video CODEC. Table 3.1 lists the formats and image sizes supported by H.261 and H.263.

Table 3.1 Video Image Format Standards

Format	Image Size	H.261	H.263
Sub-QCIF	128x96	Optional	Required
QCIF	176x144	Required	Required
CIF	352x288	Optional	Optional
4CIF	702x576	N/A	Optional
16CIF	1408x1152	N/A	Optional

Choosing a Voice Gateway Solution

There are a number of different voice gateways available for CallManager and VoIP implementations, which are divided into categories by type of gateway and the protocol running for gateway communications. The gateway selection is based on some of the following variables: analog or digital, capacity, connection type, services, features, and country of installation. Table 3.2 lists the analog VoIP gateways and the respective voice interface cards (VICs) supported. The analog gateways provide connectivity to analog phone sets, central office, and PBX. The Foreign Exchange Station (FXS) ports are used to provide dial tone for analog phones, faxes, and speakerphones, while a Foreign Exchange Office (FXO) port in a gateway is for connectivity to Central Office for analog access to the PSTN. Ear-and-mouth (E&M) ports, on the other hand, are for PBX-to-PBX signaling communication. You can determine what type of analog VoIP gateway you need by answering one of the following questions: Are FXO ports required for PSTN connectivity or is Direct Inward Dial (DID) a necessity? Another factor in selecting an analog gateway may be capacity. For example, if you require a large number of FXS ports for legacy analog connections, you would select a Cisco 3660 or a Catalyst 6000 with a 24-port FXS module rather than a smaller capacity gateway such as a VG-200 or Cisco 2600.

Table 3.2 Analog VoIP Gateways

Gateway	E&M	FXO	FXS	DID/CLID
Catalyst 4000 Access Gateway Module	Yes	Yes	Yes	12.1(5)T/12.1(5)T
Catalyst 6000 Voice T1/E1 Module	No	No	Yes	No/Yes
Cisco 1750	Yes	Yes	Yes	Future
Cisco 2600	Yes	Yes	Yes	12.1(3)T/12.1(2)XH
Cisco 3600	Yes	Yes	Yes	12.1(3)T/12.1(2)XH
Cisco 3810	Yes	Yes	Yes	12.1(3)T/12.1(2)XH
Cisco AS5300	No	No	No	N/A
Cisco 7200	No	No	No	N/A
Cisco 7500	No	No	No	N/A
Cisco DT-24+ and DE-30+	No	No	No	N/A
Cisco VG-200	H.323v2	Yes	Yes	12.1(5)XM1

If higher capacity voice channels are required to either the PSTN or PBX, a digital gateway may be more effective. Table 3.3 lists the interfaces and features supported on the various hardware platforms. The different gateways support two main signaling types: either ISDN Primary Rate Interface (PRI) or channel associated signaling (CAS) for T1 or E1. ISDN PRI, meanwhile, utilizes a "D" channel for signaling. ISDN PRI is classified as out-of-band signaling since there is a channel dedicated for signaling, whereas, CAS signaling (also referred to as robbed-bit signaling) uses some of the bandwidth from each channel. CAS signaling forms include loop start, ground start, and E&M. T1CAS supports automatic number identifier (ANI) and dialed number identification service (DNIS) as well, which are also known as Caller ID and Called Party Number. DNIS returns to the called number they dialed. Determining which PRI type of interface is required depends on whether you're connecting your gateway to a PBX or PSTN. Typically, if the gateway is connecting to a PBX, you will need a Network Side PRI interface, since the PBX is on the "user side." Normally, the PSTN (with a switch like a DMS100) functions as the "network side" and the gateway needs a User Side PRI interface. The gateway could be connected to both the PSTN and PBX with ISDN PRI configured appropriately for each interface.

Table 3.3 Digital VoIP Gateways

Gateway	T1 CAS	E1 CAS	E1/R2	User Side PRI	Network Side PRI	User Side BRI	Network Side BRI	DID/CLID
Catalyst 4000								
WS-X4604-GWY	Yes	Yes	Yes	Yes	Yes	Future	Future	Yes/Yes
Catalyst 6000								
WS-X6608-T1/E1	Future	No	No	Yes	Yes	No	No	Yes/Yes
Cisco 1750	No	No	No	No	No	12.1(5)YB1	12.1(5)YB1	
Cisco 2600	Yes	12.1(3)T	12.1(3)T	12.1(3)T	12.1(3)T	Yes	12.2(1)T	Yes/Yes
Cisco 3600	Yes	12.1(3)T	12.1(3)T	12.1(3)T	12.1(3)T	Yes	12.2(1)T	Yes/Yes
Cisco 3810	Yes	Yes	No	No	No	Yes	No	Yes
Cisco 7200	Yes	12.1(3)T	12.1(3)T	12.1(3)T	12.1(3)T	No	No	Yes/Yes
Cisco 7500	Yes	12.1(3)T	12.1(3)T	12.1(3)T	12.1(3)T	No	No	Yes/Yes
Cisco AS5300	Yes	Yes	Yes	Yes	12.0(7)T	No	No	Yes/Yes
DT-24/DE-30	No	No	No	Yes	Yes	No	No	Yes
Cisco VG-200	H.323v2	H.323v2	Yes	Yes	Yes	Yes	Yes	Yes//Yes

In choosing a gateway, attention should be given to ensure it supports three important features:

- CallManager Redundancy
- DTMF Relay
- Supplemental Services

Whether the installation is a large enterprise deployment with multiple CallManager clusters or a simple two-node cluster, it's important CallManager Redundancy is supported by the gateway of choice. CallManager Redundancy is required since an AVVID network needs to have the same high level of availability as the traditional PBX. You do not want your access to a PBX or PSTN compromised by a single point of failure. DTMF uses two frequencies, a high and low tone to distinguish numbers on a telephone keypad. This signaling is usually carried over a 64 Kbps voice circuit, and accomplished with little problem, but with lower-bit CODEC the signal can be lost or unrecognizable. The gateways provide out-of-band support for passing DTMF signals across the VoIP network via the gateway protocols, as illustrated in Table 3.4. The AVVID gateway needs to provide support for other user telephony services such as Call Hold, Call Handling, and N-way Conference. These are normal traditional voice services, which should be considered basic requirements for an AVVID network.

Table 3.4 IP Telephony Gateways and Protocols

Gateway	H.323	MGCP	Skinny
Catalyst 4000 Access Gateway Module	Yes (PSTN)	Future	Yes (Conferencing and MTP transcoding services)
Catalyst 6000 Voice T1/E1 Module	No	Future	Yes
Cisco 1750	Yes	Yes	No
Cisco 2600	Yes	Yes	No
Cisco 3600	Yes	Yes	No
Cisco 3810	Yes	Future	No
Cisco 7200	Yes	No	No
Cisco 7500	Future	No	No
Cisco AS5300	Yes	No	No
Cisco DT-24+ and DE-30+	No	No	Yes
Cisco VG-200	Yes	Yes	No

In the following sections, we will discuss the different hardware platforms in regards to what interfaces and protocols are supported as well as what application or environment is suitable.

Cisco 1750

The Cisco 1750 is an entry-level multiservice router supporting VoIP. The 1750 router can support 2 to 4 analog voice ports, as well as the FXS, FXO, and E&M voice interface cards. It is typically used as a small-scale toll bypass solution. These voice interface cards are the same modules used with the Cisco 2600 and 3600 series routers. VoIP call admission via RSVP has been supported on the 1750 since release 12.1(3)XI.

Cisco IOS Release 12.1(3)T supports the following features:

- Low Latency Queuing for Voice over Frame Relay
- H.323 voice (v2) support
- RAS protocol voice (v2) enhancement support
- DiffServ
- Fax relay enhancements

Another new feature supported in 12.1(5)XM or later is digital CAS (E&M) interfaces in addition to the analog (FXO, FXS, and E&M) interfaces. The 1750 router is not a typical gateway in a CallManager environment. Again, the Cisco 1750 is a good fit for the small office environment deploying a toll-bypass solution. For example, let's imagine we have a company with a main office in a metropolitan area and have several small field offices. In the field offices, the 1750 could be used to connect to a key system with a FXO port on the switch trunk side and have WAN connectivity back to the main office. This would allow the several field offices to utilize their current WAN circuits for voice and data, which would eliminate long distance charges back to headquarters.

Cisco 2600

The 2600 and 3600 Series routers are the mid-range multiservice platform routers. The 2600 series is comprised of ten models, three performance levels, and three topology types (listed in Table 3.5).

Table 3.5 Cisco 2600 Series Routers

Model	Interface	Processor	Performance (pps)
2610	One 10 Mbps Ethernet	40MHz	15,000
2611	Two 10 Mbps Ethernet	40MHz	15,000
2612	One 10 Mbps Ethernet One Token Ring	40MHz	15,000
2613	Two Token Ring	40MHz	15,000
2620	One 10/100 Mbps Ethernet	50MHz	25,000
2621	Two 10/100 Mbps Ethernet	50MHz	25,000
2650	One 10/100 Mbps Ethernet	80MHz	37,000
2651	Two 10/100 Mbps Ethernet	80MHz	37,000

All the 2600 Series routers provide the same number of network module slots and WAN interface slots, which is one and two, respectively. The network module slot can hold one two-slot voice/fax network module or high-density voice/fax module. The two-slot voice/fax network module allows for two of the two-port FXS, FXO, and E&M voice interface cards or any combination thereof. The high-density voice modules, meanwhile, support one to two T1 or E1 ports.

The Cisco 2600 can be used in a traditional VoIP environment providing toll-bypass, or be integrated into an AVVID architecture. The 2600 router communicates with CallManager via H.323 and MGCP protocol. As of CallManager 3.0(5), MGCP protocol is supported on FXS and FXO analog interfaces. A small company with a 25-user site could use a 2600 router with two two-port FXO modules providing four analog lines for outgoing and incoming calls. This router would also provide IP WAN connectivity. As the site grows, the Cisco 2600 can be upgraded from the two two-port FXO modules to a high-density voice module, NM-HDV-1T1-24, which will provide more trunks for the user base in this larger environment.

Cisco 3600

The 3600 Series router covers the following models 362x, 364x, and 366x. The 3600 Series routers are similar to the 2600 Series. The main differences between the 2600 and 3600 are scalability and performance. Table 3.6 lists the different models and their capacity of analog and digital channels. The Cisco 3660 Voice Gateway interoperates with H.323-compliant voice and videoconferencing applications such as Microsoft NetMeeting, as well as third-party H.323-compliant

gateways and gatekeepers. The 3660 router is a high-performance (100 mips) gateway and gatekeeper, which can be used in an AVVID CallManager deployment or in a toll bypass VoIP environment.

Table 3.6 Cisco 3600 Series Analog/Digital Scalability

Model	Slots	Analog PSTN	Digital PSTN DS0s T1/E1	Performance (pps)
Cisco 3620	2	4	48/60	20–40,000
Cisco 3640	4	12	136/180	50–70,000
Cisco 3660	6	24	288/360	120,000

VG-200

The VG-200 is a VoIP gateway which is modular, 1U rack-mounted, and based on the Cisco 2600 chassis. The VG-200 is similar to the Cisco 2600, but without routing capabilities. It supports connectivity to PBX, PSTN, analog and digital dial access, and legacy voice mail systems, and has the capacity for up to four analog ports or two digital ports. A common configuration of a small branch office is the VG-200 with a NM-2V module and two VIC cards. The VIC cards can be FXS, FXO, E&M, or ISDN Basic Rate Interface (BRI). An upgraded configuration would have one digital T1/E1 voice module (NM-HDV), which has either 1 or 2 ports. This configuration supports up to 60 voice channels. The VG-200 communicates with Cisco CallManager via MGCP and H.323 protocol. FXS and FXO are the only ports supported under the MGCP configuration mode with CallManager. The next section contains an example configuration of how you might implement MGCP in a small office environment using a VG200. This example illustrates the use of multiple FXO and FXS ports. The FXS ports could be used for a legacy conference speakerphone and fax machine with the FXO ports providing PSTN connectivity.

Configuring and Installing a VG200 with MGCP

The following is the step-by-step procedure to configure a VG200 with MGCP as its gateway protocol.

1. First, configure an IP address on the VG200's Ethernet interface and enable the interface:

```
router(config)#interface fastethernet 0/0
router(config-if)#ip address 10.7.1.252 255.255.255.0
router(config-if)#no shut
```

2. Next, assign a unique name to the VG200 so the CallManager server can identify it:

```
router(config)#hostname DNVRVG200a
```

This is how CallManager keeps track of the MGCP network devices it is communicating with. This name must be unique.

3. Configure the VG200 to run MGCP as a signaling protocol.

```
DNVRVG200A(config)#mgcp
```

4. Configure the IP address (or DNS Name) for the CallManager server:

```
VG200(config)#mgcp call-agent 10.7.1.1
```

5. Select the CODEC type and the DTMF relay function:

```
DNVRVG200A(config)#mgcp dtmf-relay codec all mode out-of-band
```

6. To enable support for Cisco CallManager within MGCP, enter the following command:

```
DNVRVG200A(config)#ccm-manager mgcp
```

NOTE

Use the command **show voice port** to determine the type of ports the VG200 has and which order they are installed in.

7. Bind the MGCP application to the voice ports.

- FXO Port:

```
DNVRVG200A(config)#dial-peer voice 1 pots
DNVRVG200A(config)#application MGCPAPP
DNVRVG200A(config)#port 1/0/0
```

- FXO Port:

```
DNVRVG200A(config)#dial-peer voice 2 pots
```

```
DNVRVG200A(config)#application MGCPAPP
DNVRVG200A(config)#port 1/0/1
```

- FXS Port:

```
DNVRVG200A(config)#dial-peer voice 3 pots
DNVRVG200A(config)#application MGCPAPP
DNVRVG200A(config)#port 1/1/0
```

- FXS Port:

```
DNVRVG200A(config)#dial-peer voice 4 pots
DNVRVG200A(config)#application MGCPAPP
DNVRVG200A(config)#port 1/1/1
```

NOTE

In some Cisco IOS versions, the application **MGCPAPP** command is case-sensitive. As a precaution, always enter this command in uppercase. Make certain the voice ports are enabled as well. Executing the command **no shut** enables both ports on a VIC.

- FXO Port:

```
DNVRVG200A(config)#voice-port 1/0/0
DNVRVG200A(config-voiceport)#no shut
```

- FXS Port:

```
DNVRVG200A(config)#voice-port 1/1/0
DNVRVG200A(config-voiceport)#no shut
```

- Under Dial Peers:

```
destination-pattern
session-target
```

- Under Voice Ports:

```
connection { plar | tie line | trunk }
```

WARNING

The commands listed for Under Dial Peers and Under Voice Ports should not be configured in the MGCP gateway for MGCP-managed endpoints (those with application MGCAPP command in their dial-peer statement). It will cause communication problems with your CallManager.

8. The VG200 is configured to communicate with the CallManager server. It will periodically send out messages attempting to establish a connection. When the CallManager server configuration is complete, the connection should automatically establish itself.
9. The following is the complete configuration for the VG200 router, DNVRVG200A, for this document:

```
DNVRVG200A#show running-config
Building configuration...

Current configuration : 1244 bytes
!
version 12.1
no service single-slot-reload-enable
no service pad
service timestamps debug uptime
service timestamps log uptime
no service password-encryption
!
hostname DNVRVG200A
!
boot system flash
boot system rom
boot system tftp vg200 10.7.1.253
no logging buffered
logging rate-limit console 10 except errors
no logging console
enable secret jUie9834/#@skui
enable password #####
```

```
!
ip subnet-zero
no ip finger
no ip domain-lookup
!
mgcp
mgcp call-agent 10.7.1.1
mgcp dtmf-relay codec all mode out-of-band
mgcp sdp simple
call rsvp-sync
!
!
!
ccm-manager mgcp
!
!
interface FastEthernet0/0
 ip address 10.7.1.252 255.255.0.0
 no ip mroute-cache
 speed auto
 full-duplex
!
ip default-gateway 10.7.1.254
ip classless
no ip http server
!
snmp-server engineID local 000000090200000196983000
snmp-server community public RO
!
voice-port 1/0/0
!
voice-port 1/0/1
!
voice-port 1/1/0
!
```

```
voice-port 1/1/1
!
dial-peer voice 1 pots
 application mgcpapp
 port 1/0/0
!
dial-peer voice 2 pots
 application mgcpapp
 port 1/0/1
!
dial-peer voice 3 pots
 application mgcpapp
 port 1/1/0
!
dial-peer voice 4 pots
 application mgcpapp
 port 1/1/1
!
!
line con 0
 transport input none
line aux 0
line vty 0 4
 password ww
 login
!
end

DNVRVG200A#
```

Cisco MC3810

The Cisco MC3810 integrates data, voice, and video applications into a single box solution. The MC3810 supports voice connectivity using the following methods: Voice over Frame Relay (VoFR), Voice over ATM (VoATM), and VoIP. The interfaces supported by the 3810 are the following:

- Six analog ports (FXO, FXS, and E&M)
- Digital T1/E1 (Drop and Insert, CAS, CCS, and PRI QSIG)
- Ethernet 10Base-T
- Two serial ports
- Four BRI voice ports

The analog interfaces are not compatible or interchangeable with Cisco's 1750, 2600, and 3600 routers. The MC3810 can interface with the PSTN or PBX via digital connection. The digital T1 PBX connection supports 24 channels for voice with the following compression CODECs: G.723.1, G.729, G.729a, and G.726. This router provides similar Voice over X functionality as these routers, in addition it has video capability such as circuit emulation over ATM, and H.323 gatekeeper for Video over IP. The MC3810 video features allow organizations to get rid of H.320 ISDN dial-up circuits.

Cisco 7200/7500

For enterprises seeking a high capacity and performance VoIP solution, the 7200 series routers are a viable choice. The Cisco 7200 with six slots can be equipped with up to 20 T1s or 18 E1s, supporting voice-over-packet applications. The DS0 channel from a T1/E1 is switched into the Digital Signal Processor (DSP) to perform the TDM-to-packet conversion of the bearer information present on a DS0, while the Cisco 7200 router supports two types of port adapters, one with DSPs and one without. The adapters without DSPs can use DSPs from the other DSP-capable adapters. The PA-MCX-2/4/8 TE1 is a non-DSP adapter and works with the PA-VXx-2TE1+ type adapters. The PA-VXx-2TE1+ adapters, on the other hand, provide up to two T1 or E1 interfaces. Table 3.7 lists the number of voice channels supported, based on the CODEC complexity.

Table 3.7 7200 T1/E1 Voice Port Adapters

Product	High-Complexity Voice Channels	Medium-Complexity Voice Channels
PA-VXB-2TE1+	24	48
PA-VXC-2TE1+	60	120

When installing multiple T1 or E1 voice port adapters, you must use a combination of the PA-VXB-XXX and PA-MCX-XXX adapters in order to share the DSP resources. Table 3.8 lists the Cisco IOS release supported on the given platform for the given digital signaling protocol.

Table 3.8 7200/7500 Signaling Protocols

Signaling Protocol	7200	7500
CAS T1	12.0(5)XE3, 12.0(7)XK, 12.1(1)T*	12.1(3)T
CAS E1	12.0(5)XE3*	12.1(3)T
Q.SIG	12.0(7)XK*, 12.1(2)T	12.1(3)T
PRI Q.931 User Side	12.1(3)T	12.1(3)T
PRI Q.931 Network Side	12.1(3)T	12.1(3)T
R2 Signaling	12.1(3)T	12.1(3)T
Transparent CCS	12.1(3)T	12.1(3)T
Feature Group D	12.1(4)T	12.1(4)T
Multi-D Channel	12.1(3)T	12.1(3)T
RAI	Future	Future

NOTE

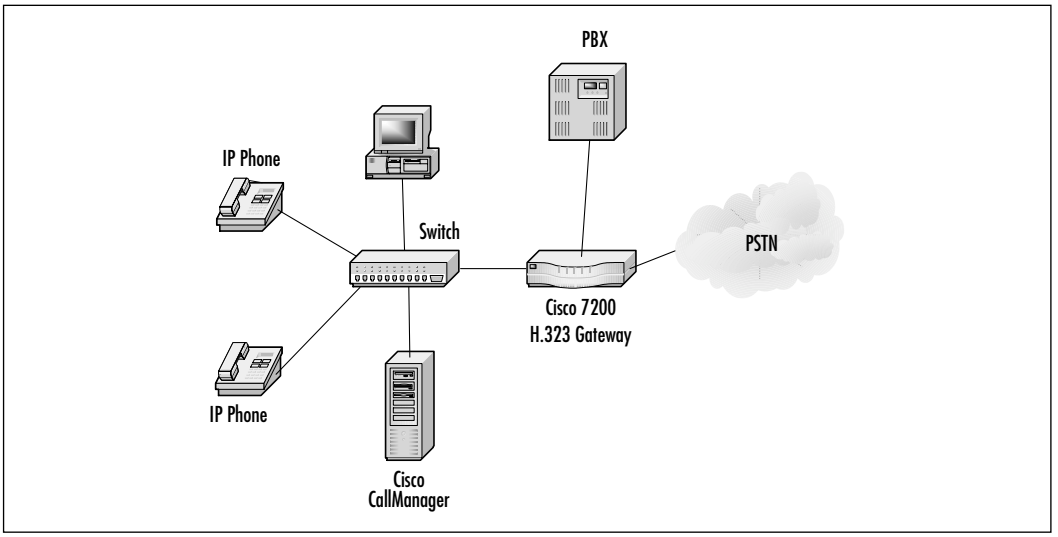
IOS Releases 12.1(3)T and earlier are supported on digital voice port adapters PA-VXC-2TE1, PA-VXB-2TE1. IOS Releases 12.1(2)T and later are supported on the enhanced digital voice port adapters PA-VXC-2TE1+ and PA-VXB-2TE1+.

The 7200 can perform as a high density H.323 AVVID gateway with a connection to the PSTN and PBX, providing digital connection to both the PSTN and PBX (as illustrated in Figure 3.1).

Cisco AS5300/AS5800

The AS5300 is a H.323-compliant enterprise-based VoIP gateway solution. The AS5300 can scale to 96/120 connections based on the T1 or E1 modules installed. This is accomplished with a quad T1 or E1 modules and two voice feature cards that support 48/60 voice connections per card.

Figure 3.1 Cisco 7200



The Cisco AS5800 supports up to 1344 voice ports in a single system, thus offering a high concentration of VoIP DSPs in a single voice gateway. The Cisco AS5800 supports two trunk cards: the 12-port T1/E1 and the channelized T3 termination card. The channelized T3 card provides 672 trunks, with a maximum of two cards permitting 1344 trunks per AS5800 chassis. It supports voice feature cards with a port density of 96 to 336 ports per card, as shown in Table 3.9.

Table 3.9 AS5800 Voice Feature Cards

Part Number	Description
DS58-336-MC-VOx	AS5800 336-Port Medium Complexity Voice Card
DS58-192-MC-VOx	AS5800 192-Port Medium Complexity Voice Card
DS58-192VOx	AS5800 192-Port Voice Card
DS58-96VOx	AS5800 96-Port Voice Card

Cisco DT-24+/DE-30+

The DT-24/DE-30+ is a PCI-based digital gateway card that supports up to 23 or 30 voice channels. The card can be installed in a Cisco CallManager server or any other PC with PCI slots, and only draws power from the PC. The DT-24/DE-30+ card provides connectivity to the PSTN or a PBX. The DT-24/DE-30

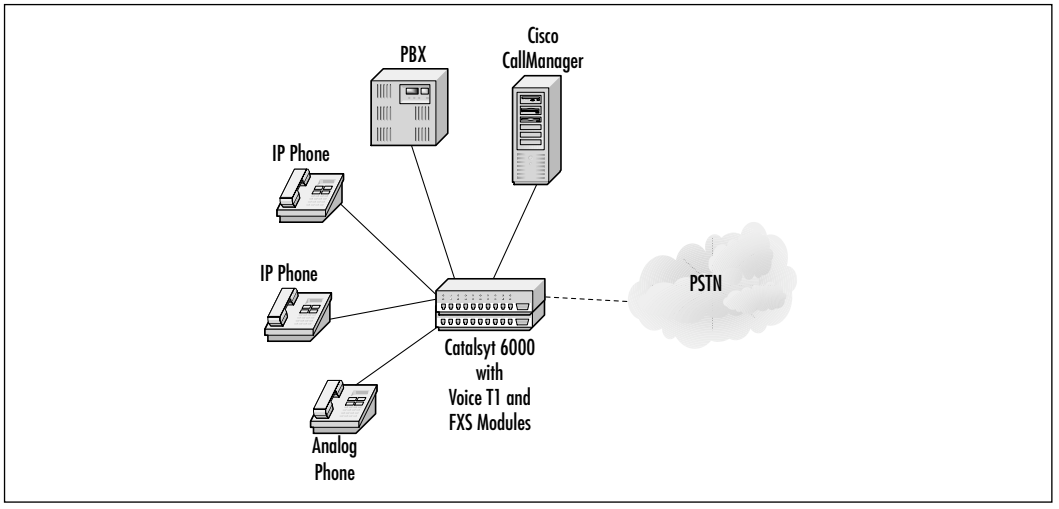
communicates with Cisco CallManager via Skinny Station Protocol and supports supplementary services such as hold, transfer, and call forwarding. The DT-24/DE-30 features ISDN PRI (T1/E1) rates for its trunk interface and a 10Base-T Ethernet port for the IP network. If more trunks are needed, you can install as many of the DT-24 or DE-30 cards that any given PC with PCI slots will support.

These cards provide similar functionality as the NM-HDV-1T1-24 or NM-HDV-1E1-30 modules for the Cisco 2600 and 3600 Series routers. If your company has not invested in a router such as a 2600 or 3600 and has available PCs with PCI slots, the DT-24 could be utilized to provide ISDN PRI connectivity. I foresee the majority of development in further software and hardware enhancements in the router and switch class gateways. This will allow enterprises to leverage their current investment and knowledge in these platforms.

Catalyst 6000

The Catalyst 6000 is an enterprise-class voice-capable switch, capable of supporting analog and digital voice interfaces. It is a highly scalable switch, which makes it an integral component of an AVVID network. As illustrated in Figure 3.2, the Catalyst 6000 can offer connectivity to the PSTN, legacy PBX, analog phones, and IP phones. It provides inline power to Cisco IP phones via a 48-port 10/100 Ethernet module, WS-X6348-RJ45V.

Figure 3.2 Catalyst 6000



It provides connectivity to legacy analog devices via a 24-port FXS module, WS-X6624-FXS, which can be used for analog phones, faxes, speakerphones, and modems. The analog FXS module acts as the gateway between the analog devices and the IP network. By allowing for analog gateway functionality, organizations can extend the useful life of their legacy analog devices. It also helps in the migration to an IP telephony network by supporting existing fax machines and conference speakerphones. Cisco does have an IP-based speakerphone, Cisco 7935, which was co-developed with Polycom. The analog to IP communication is achieved via the voice CODECs of either G.711 or G.729a.

Digital PSTN and legacy PBX access is achieved with a Catalyst 6000 T1 or E1 voice module, which is designed for larger enterprise campus environments. The signaling supported for PSTN connections are Common channel signaling (CCS) and ISDN PRI. The T1 module supports 23 channels in either signaling mode, while the E1 module supports 29 channels for CCS and 30 for ISDN PRI mode. The modules can also be configured to support transcoding and conference bridging by configuring some ports for PBX or PSTN connectivity and others for use as DSP resources.

Catalyst 4000

The Catalyst 4000 series is comprised of the 4003, 4006, 4908G, and 4912G models. The Catalyst 4003 and 4006 are modular chassis-based switches, which are a large part of the Cisco AVVID architecture. The Catalyst 4000 series family (which is actually a scaled down version of the Catalyst 6000) is targeted at branch offices, enterprise wiring closets, and mid-range organizations. The Catalyst 4006 supports inline-power Ethernet modules for Cisco IP phones while the Catalyst 4003 uses inline-power patch panels. It must use an external Auxiliary DC Power Shelf to provide the needed power to the IP phones. For providing gateway services to the IP telephony network, Cisco has an Access Gateway Module (AGM), WS-X4604-GWY, for the Catalyst 4000.

The AGM allows the Catalyst 4000 to be an integrated solution providing IP WAN routing, gateway functionality to the PSTN and PBX, and DSP resources for CallManager. It is supported in both the 4003 and 4006, and uses the same VIC and voice WAN interface card (VWIC) modules as the 1750, 2600, and 3600 series routers. The AGM has one dedicated VIC slot and two VWIC slots, which holds either a VIC, VWIC, or WIC module. The AGM can be connected to the PSTN or a PBX, and act as a H.323 gateway for CallManager. Analog devices such as phones, speakerphones, and faxes can be connected to the AGM via an 8-port RJ-21 FXS module, WS-U4604-8FXS, or one of the VIC card ports. The

8-port FXS module is installed in the FlexSlot on the AGM, and the FlexSlot supplements the other three VIC slots. In the future, the AGM will support a 16-port FXS module allowing the AGM to support up to a total of 22 analog ports. The three VIC slots, each holding a two-port card, provide the six other ports. FXS, FXO, or E&M VIC modules can be used in these three VIC slots.

Catalyst 4224

One of the newer switches to the Cisco line is the Catalyst 4200 Series family. The Catalyst 4200 Access Gateway series switches allow voice, video, and data to be offered to the small branch environment looking to deploy IP telephony solutions. The Catalyst 4224 is a 2U rack mounted switch with 24 10/100 Ethernet port and modular slots for voice and WAN modules. The same VIC and VWIC modules mentioned previously for the Catalyst 4000 Series can be used in the Catalyst 4224.

The 4200 Series was designed for small branch offices to deploy a complete IP telephony solution, as well as enterprises in a centralized call processing CallManager model. With a centralized call processing CallManager model if a remote office loses its connection to the central office where the CallManager is located, they will be unable to perform any voice calls from that office unless they have a backup link to the central office or have Survivable Remote Site (SRS) telephony software. If the company has several remote offices, it could be rather expensive to have backup links for all these sites.

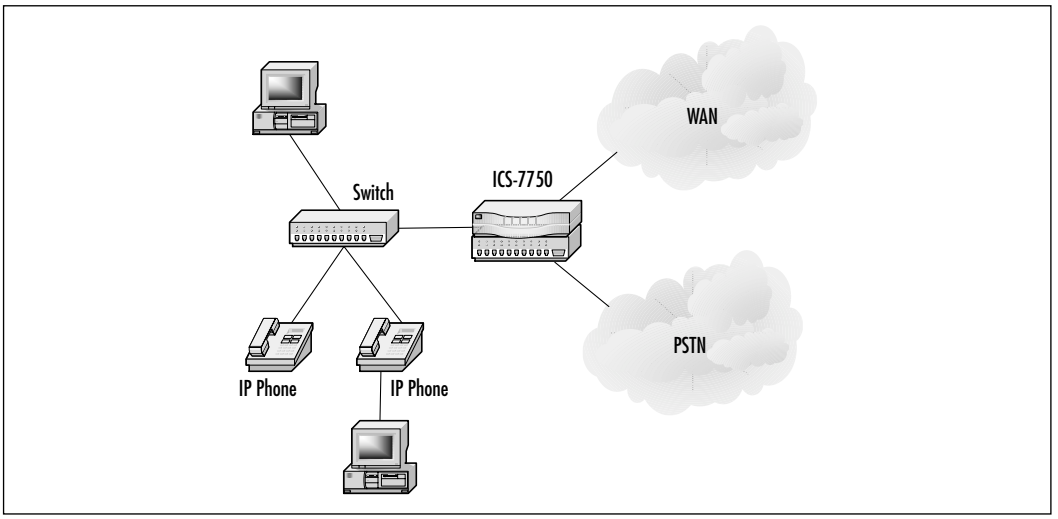
To provide a more cost-effective solution, Cisco developed SRS, which is part of the IOS software that runs on a 4224 switch. As of this writing, SRS is also available on Cisco 2600 and 3600 routers. This technology will also be included in other Cisco products, such as the Cisco 175x router, and the Catalyst 4000 AGM by Q4 2001. SRS automatically detects a network failure, and using the Cisco Simple Network Automated Provisioning (SNAP) capability, reconfigures the router to provide call processing for the IP phones in that location. When the WAN is restored, the router will shift call-processing functions back to the CallManager cluster.

Smaller routers such as the 1750, 2600, and 3620 along with the Catalyst 4224 will support up to 24 phones. Meanwhile, the Cisco 3640 and Catalyst 4000 AGM support up to 48 phones, and the Cisco 3660 supports up to 144 phones. Up to two lines per phone are supported per system.

ICS 7750

The Integrated Communication System (ICS) 7750 combines the features of a multiservice router, a voice gateway, and call processing into a single chassis-based solution. The ICS 7750 is not a typical VoIP gateway, but it's covered because the gateway module utilizes the same interface cards as some of the IOS-based gateways discussed in this chapter. The ICS 7750 is targeted to branch offices and midrange organizations—a typical installation supporting up to 150 users. A complete IP telephony solution can be deployed with an ICS 7750, Catalyst 3524XL-PWR switch, and Cisco IP phones (as shown in Figure 3.3).

Figure 3.3 Cisco ICS-7750



The System Processing Engine (SPE) 200 is the call-processing component, which is a CallManager server on a blade. The module has an Intel Pentium II CPU, 512MB of RAM, a 6.4GB hard drive, as well as Windows 2000 with SQL Server and CallManager.

The Multiservice Route Processor (MRP) 200 is a voice-and-data-capable router that can carry voice and data traffic over an IP network and can link small-to-medium-size remote Ethernet LANs to central-offices LANs over different types of WAN links. The MRP utilizes the same VIC, WIC, and VWIC modules as the 1750, 2600, and 3600 series routers. The MRP 200 has a capacity of two T1 ports for voice, as well as one port for data.

G.711 and G.729a are the voice CODECs used to support communication between multiple IP and analog devices within campus and WAN environments.

These CODECs are achieved by the Packet Voice/fax DSP module (PVDM) in the MRP. The MRP 200 has the capacity for 40-channel PVDM modules.

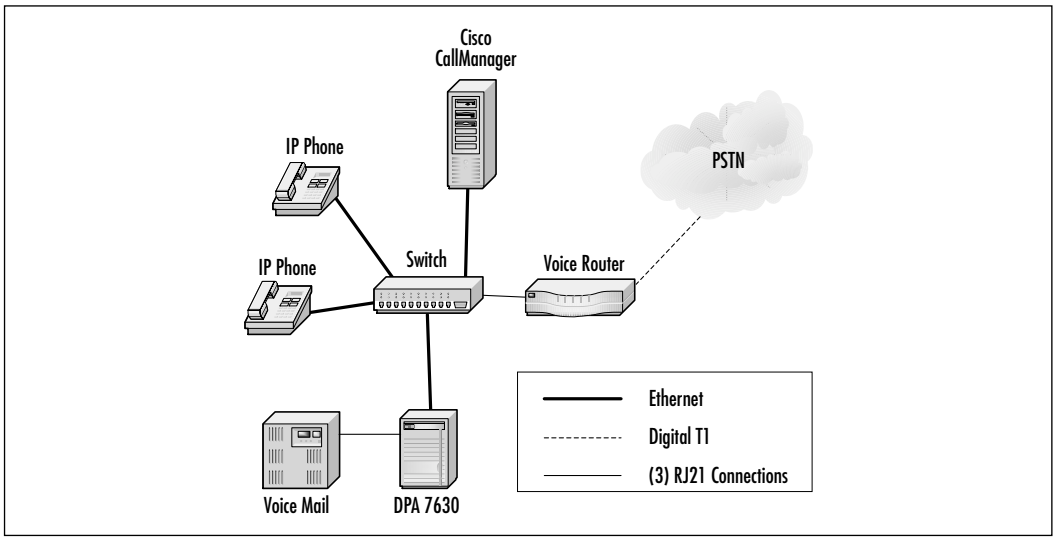
NOTE

Although the MRP 200 supports two T1 ports, it is limited to 24 simultaneous calls.

DPA 7610/7630 Voice Mail Gateway

Another component of an AVVID solution is the DPA 7610/7630 Voice Mail Gateway. The Digital PBX Adapter (DPA) 7610 and 7630 are VoIP gateways that allow legacy voice mail systems to communicate with a Cisco CallManager. The DPA 7610 and 7630 have a 10/100 Ethernet port and one or three RJ-21 ports, respectively. The DPA uses the RJ-21 ports, which provide 8 to 24 four-wire digital lines to interface with the legacy voice mail system. It communicates with the CallManager through the Skinny Station protocol via the 10/100 Ethernet port, and emulates an IP phone in order to communicate with the CallManager. The DPA then allows simultaneous communication with a CallManager and legacy PBX to the voice mail system (as illustrated in Figure 3.4).

Figure 3.4 DPA 7630 Voice Mail Gateway



The DPA supports G.711 and G.729a voice CODECs to communicate with the voice mail system, as well as the Avaya (formerly Lucent) Definity and Meridian PBX along with the Octel voice mail system. Incoming/outgoing messages and message waiting indicator (MWI) commands between CallManager and the Octel voice mail are done through the DPA.

Choosing a Video Gateway Solution

Cisco's IP/VC 3500 family of products satisfies the video part of their AVVID multiservice architecture. The IP/VC family covers video conferencing solutions from the lower-end desktop to the high-end conferencing room implementation. Considering today's business environment and the never-ending effort to curb expenses, companies are looking at more cost-effective ways of conducting business meetings and presentations. One way is to institute a video conferencing solution. In the following sections, we'll discuss some of the products Cisco offers in this area, including integrating with legacy video conferencing solutions.

IP/VC 3510 MCU

The IP/VC 3510 MCU merges three or more H.323 videoconference endpoints into a single multiparticipant meeting (as shown in Figure 3.5). The 3510 MCU is able to maintain ad hoc and scheduled videoconferences, and each unit can support up to 15 sessions at 128 Kbps, nine sessions at 384 Kbps, seven sessions at 512 Kbps, five sessions at 768 Kbps, or three sessions at 1.5 Mbps. The unit can support multiple conferences with a limiting factor of 15 sessions. Participants can join through the Web interface, or by having the MCU dial to them. The MCU can be cascaded together to support more sessions. In fact, the conference capacity for multiple MCUs is 48 sessions. Videoconferencing, meanwhile, with T.120 data sharing, is available along with audio only calls.

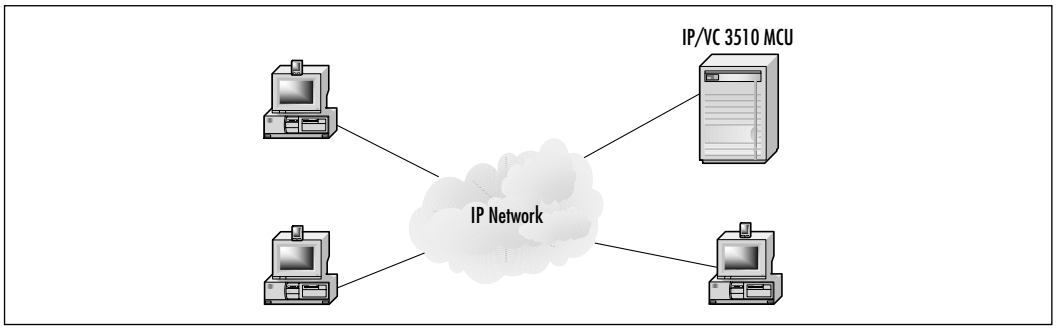
IP/VC 3520 and 3525 Gateway

The IP/VC 3520 and 3525 provide translation services between H.320 and H.323 networks. This allows companies to connect legacy ISDN H.320 videoconferencing systems to IP-based H.323 networks, letting users conduct videoconferencing across the IP LAN or via the PSTN. The IP/VC 3520 is available in five different hardware configurations with the following features:

- Four BRI interface ports
- Four V.35 interface ports

- Combination of two BRI interface ports and two V.35 ports
- Audio CODEC transcoding between G.711 and G.728
- Audio CODEC transcoding between G.711 and G.723
- Channel bonding

Figure 3.5 IP/VC 3510 MCU



The Cisco IP/VC 3520 Gateway can bond to a maximum of three ports for calls that require transfer rates of up to 384 Kbps. Each BRI interface allows a 128 Kbps call. Units with four BRI interface ports can simultaneously support four 128 Kbps calls, two 256 Kbps calls, or one 384 Kbps call and one 128 Kbps call. Each V.35 port supports transfer rates from 56 Kbps to 768 Kbps. The V.35 ports either RS-366 or V.25bis signaling. Both the 3520 and 3525 have an embedded gatekeeper, where each supports video formats of H.261 and H.263. The main differences between the two models are their interface connections and the scalability of calls.

The 3525 gateway supports ISDN PRI T1 and E1 connections. The IP/VC 3525 T1 model supports up to three videoconferencing calls at 384 Kbps, while the 3525 E1 model supports five calls at 384 Kbps. If less quality is acceptable, 13 calls at 128 Kbps is supported. Keep in mind, the 3525 supports the same audio transcoding formats as the 3520 (Figure 3.6).

IP/VC 3530 VTA

The IP/VC 3530 Video Terminal Adapter (VTA) is a 1U rack mounted unit with one Ethernet port and two V.35 ports. The 3530 VTA (shown in Figure 3.7) connects a single H.320 system to an IP network, and translates from a H.320 ISDN-based network to a H.323 IP-based network. The video session can perform at 128 Kbps and up to 768 Kbps across the IP network, connecting to a multipoint

conference hosted by an IP/VC 3510 MCU by going through an IP/VC 3520 or 3525 gateway. Like all the other IP/VC products, it supports T.120 for data collaboration. The encoding formats supported for video are H.261 and QCF/CIF, while those supported for audio are G.711, G.722, and G.728.

Figure 3.6 IP/VC 3520/3525 Gateway

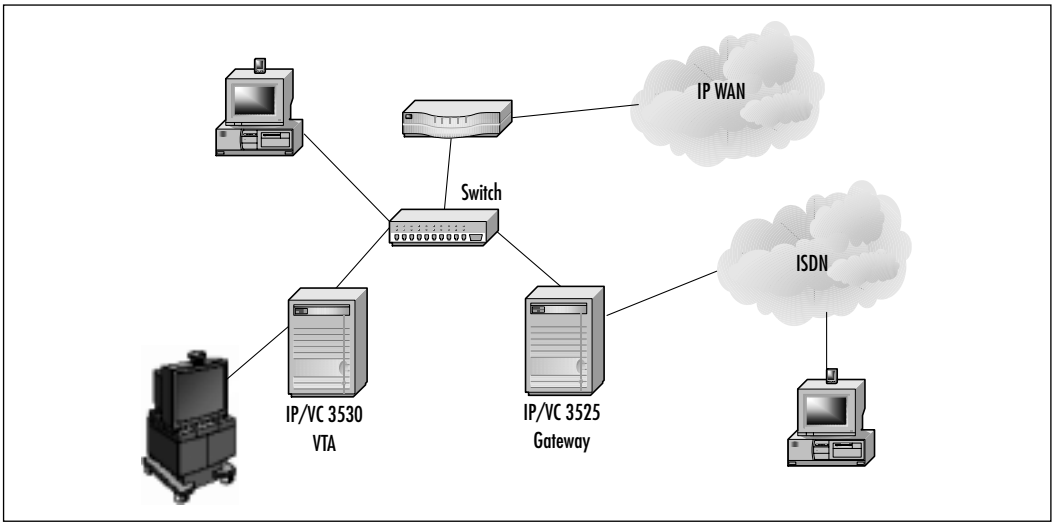
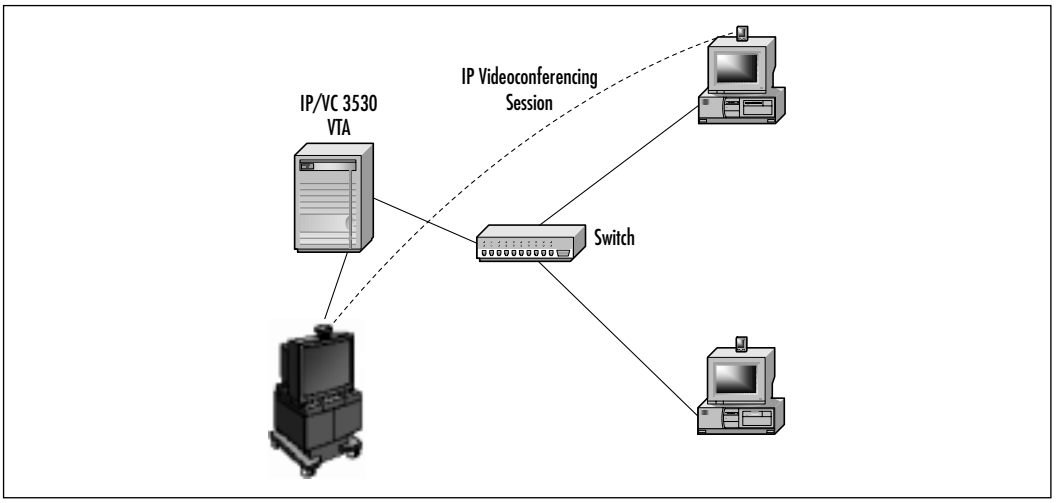


Figure 3.7 IP/VC 3530 VTA



IP/VC 3540

The IP/VC 3540 videoconferencing system is a multipoint conferencing, gateway, and data collaboration integrated solution. The IP/VC 3540 solution includes an IP/VC 3544 chassis and 3540 modules. The IP/VC 3544 chassis is a 2U 19 inch rack mountable unit that has four slots in a Compact PCI (cPCI) bus backplane. The IP/VC 3540 system module manages the cPCI bus, and the chassis supports two modules, an IP/VC 3540 MCU, Application Server, and H.320-to-H.323 Gateway module. The management is performed through a Web interface by java-enabled Web browsers.

The IP/VC 3540 MCU module supplies real-time voice, videoconferencing, and T.120 data collaboration capabilities for companies desiring high quality and scalability. The multipoint conferences can be scheduled or ad-hoc, while the quality of video sessions range from 768 Kbps for high quality to 2 Mbps for super-quality. The IP/VC 3540 MCU module supports multipoint videoconferences with up to 100 participants, and comes in the following options: 30-, 60-, and 100-sessions (128 Kbps). The following is a list of user limitations with performance ratings:

- **100-session MCU module** 100 participants at 128 Kbps, 50 participants at 384 Kbps, 25 participants at 768 Kbps, 10 participants at 1.5/2.0 Mbps, or 150 voice-only participants.
- **60-session MCU module** 60 participants at 128 Kbps, 30 participants at 384 Kbps, 15 participants at 768 Kbps, 5 participants at 1.5/2.0 Mbps, or 150 voice-only participants.
- **30-session MCU module** 30 participants at 128 Kbps, 15 participants at 384 Kbps, 9 participants at 768 Kbps, 3 participants at 1.5/2.0 Mbps, or 45 voice-only participants.

Based on an Intel Pentium server running Windows NT, the IP/VC 3540 Application Server (AS) acts as a T.120 data collaboration conferencing server allowing end users to perform slide presentations, whiteboard, and other applications during a conference call. The 3540 AS has the same participant user limits as the MCU module.

The IP/VC 3540 H.32-to-H.323 Gateway module provides ISDN H.320 to IP-based H.323 translation. This is the same functionality provided by the 3520 and 3525, and has two PRI ports that can be configured at T1 or E1 speeds. Along with other IP/VC products, it supports H.261 and H.263 for video format encoding. Voice encoding, meanwhile, is performed with G.711, G.722, and G.728.

Multimedia Conference Manager Services

The Multimedia Conference Manager (MCM) is part of the Cisco IOS for the following router platforms: 2500, 2600, 3600, 3810, and 7200. It services a H.323 gatekeeper and proxy, and works in conjunction with Cisco's IP/VC products. Since MCM is part of the IOS and runs on Cisco routers, it is able to scale based on organizational needs, while its gatekeeper service provides admission control, bandwidth management, address resolution, AAA services, and call routing. The proxy service of MCM ensures quality videoconferencing by providing QoS capabilities.

The gatekeeper is the control point to the H.323 endpoint, MCU, H.320-to-H.323, and VoIP gateways, as well as the IP telephony devices. The following are some of the gatekeeper's responsibilities:

- H.323 administrative zone establishment
- AAA services
- Endpoint registration via RAS
- Intra-zone and Inter-zone call establishment
- Bandwidth management
- Session management
- Accounting and billing

The proxy's job is to terminate H.323 calls from the local zone and establish communication with endpoints in another zone. Based on this intervening of the proxy, quality of service policies can be established regarding inter-zone communication. The proxy's responsibilities are as follows:

- H.323 calling signaling termination and generation
- Videoconference traffic identification
- QoS classification:
  - IP Precedence
  - RSVP
- WAN bandwidth management
- Security

Table 3.10 describes the scalability of MCM solutions based on their hardware platform.

Table 3.10 MCM Performance

Platform	IP Routing Packet per Second	H.323 Endpoints	Video Calls	Video Proxy Sessions
Cisco 7200	50–100K	3000	500	50 at 768 Kbps 75 at 384 Kbps 100 at 128 Kbps
Cisco 3660	25–100K	1800	250	25 at 768 Kbps 35 at 384 Kbps 50 at 128 Kbps
Cisco 3640	10–40K	1800	150	10 at 768 Kbps 15 at 384 Kbps 30 at 128 Kbps
Cisco 3620	10–15K	1800	75	10 at 768 Kbps 15 at 384 Kbps 30 at 128 Kbps
Cisco 262x	5–10K	900	60	2 at 768 Kbps 4 at 384 Kbps 6 at 128 Kbps
Cisco 261x	2–5K	900	60	2 at 768 Kbps 4 at 384 Kbps 6 at 128 Kbps
Cisco 3810	2–5K	900	60	2 at 768 Kbps 4 at 384 Kbps 6 at 128 Kbps
Cisco 25Xx	N/A	600	30	2 at 768 Kbps 4 at 384 Kbps 10 at 128 Kbps

Developing & Deploying…

Gateway Selection Questions

When designing and deploying a Cisco AVVID solution, the selection of a gateway for the environment will most likely be based on several criteria. The following is a sample list of questions regarding required features that should be asked prior to selecting a gateway:

- Is an analog or digital gateway required?
- What is the required capacity of the gateway?
- Is it a standalone or IOS-based gateway?
- Do you want an integrated all-in-one solution?
- Is IP routing required?
- What type of connection is the gateway going to use (analog POTS, PSTN, or PBX connection)?
- What types of supplementary services are desired?
- Is voice compression a part of the design? If so, which types?
- Is direct inward dialing (DID) required?
- Is calling line ID (CLID) needed?
- Is fax relay needed?
- What type of network management interface is preferred?
- To which country will the hardware be shipped?
- Is rack space available for all needed gateways, routers, and switches?

The need to verify equipment features and capabilities is constant as Cisco continues to update the IOS software features and routers/gateways listed in this chapter.

Summary

The importance of gateway selection is not to be overlooked, whether your emphasis is on analog or digital protocols or both. Completely understanding all the equipment's features and benefits as well as the protocols should help make this important decision easier to make.

AVVID voice gateways include standalone, IOS-based, and Catalyst switches. The gateway protocols supported are H.323, MGCP, and Skinny with SIP gaining ongoing popularity. The voice gateways range from small analog routers such as the 1750 to large scalable digital T1/E1 7200 routers and everywhere in between. The gateways can be more traditional VoIP toll bypass or total integrated all-in-one solutions like the Catalyst 4224.

For small- to medium-sized organizations, the best solution may be either a Cisco 2600 router or the Catalyst 4224. Either solution should not only be able to handle VoIP solutions but other AVVID gateway requirements as well. The 2600 Series also has expansion capabilities to help with organizational growth. However, if you do not require routing capabilities, you might look to the VG200 to provide similar solutions. When looking at the needs of a medium to large organization, one would have to look at the 3600 Series router, which provides the scalability necessary to handle the needs of a large enterprise environment. The 3660 router has the ability to support up to 12 T1, which would consequently support 2000+ users in a PSTN gateway scenario. The MC3810 would provide a one step solution for data, voice and video needs. It provides VoFR, VoIP but also VoATM. However, the MC3810 does not have the modular flexibility of the 2600 or 3600 routers. It also does not integrate with CallManager.

When you are looking for switch-based solutions with similar functionality, as the 2600/3600 Series routers do, the Catalyst 4000 would be a good choice. It supports the same modules as the 2600/3600 except for the high-density voice module, NM-HDV-XXX. For large organizations seeking high capacity and performance, the choice could be the 7200 or 7500 routers with the ability to support up to 20 T1s or 18 E1s via T1/E1 CAS or PRI signaling. A large-scale switch-based solution would be the Catalyst 6500 series utilizing the 8-port T1/E1 voice module. Since the release of the Catalyst 6513, which has 13 slots, it could theoretically scale up to 96 T1 ports providing 2300+ voice channels. Most likely the configuration would allocate some of the ports as T1 and others as DSP resources, which will be discussed in Chapter 6.

The IP/VC 3500 videoconferencing products round out the gateways for the AVVID architecture. They cover multipoint conference units, gateways, and

terminal adapters. The IP/VC provides solutions in multipoint conferencing, H.320 to H.323 translation, and legacy H.320 connectivity, while the MCM completes the videoconferencing solution by providing the gatekeeper functionality required for Video over IP.

The IP/VC 3510 MCU is a multiparticipant video and data conferencing solution, whereas the IP/VC 3520 & 3525 gateways are used as ISDN H.320 to IP H.323 gateways. The main difference between the latter two models is that the 3520 supports V.35 and ISDN BRI interfaces while ISDN PRI is available on the 3525. The IP/VC 3510 connects ISDN-based H.320 systems like a PictureTel Venue 2000 to the IP-based H.323 network. Another MCU unit provided by Cisco is the IP/VC 3540, which is a highly scalable multiparticipant videoconferencing solution. The 3540 is targeted toward the large enterprise environments, whereas, the 3510 is targeted to the low-end market. The IP/VC 3540 family of products consists of a 3544 chassis, system module, MCU module, and application server (AS) module, while the 3544 chassis has four slots with one required for a system module and three designated for the other modules. The 3540 MCU is available in three models: 30-, 60-, and 100-user types for system and nonsystems modules. The MCM is part of Cisco IOS software, which runs on the Cisco 2500, 2600, 3600, 3810, and 7200 Series routers. The MCM function is to serve as the videoconferencing gatekeeper and proxy.

Considering what's been discussed in this chapter, you should now have a greater understanding of the role gateway selection will have in developing your ongoing enterprise solutions strategies, whether the importance lay in voice, video, or both.

Solutions Fast Track

Introduction to AVVID Gateways

- ☑ In the Cisco AVVID world, there are voice and video gateways to provide connectivity to legacy networks. Cisco has voice gateways, which are standalone routers, IOS-based routers, and Catalyst switch-based routers.
- ☑ The standalone gateways include the DT-24+, DE-30+, and VG200. Router IOS-based gateway solutions are the 175x, 2600, 3600, 3810, 5300, 7200, and 7500. The switch-based gateways are the Catalyst 4000, 4200, and 6000 Series. These gateways run the following protocols: H.323, MGCP, Skinny, and SIP.

- ☑ The IP/VC 3500 family is the videoconferencing gateway products from Cisco.

Understanding the Capabilities of Gateway Protocols

- ☑ H.323 is the most supported gateway protocol, backed by the Cisco 1750, 2600, 3600, AS5300, 7200, and 7500 Series routers.
- ☑ Skinny Station Protocol allows a Skinny client to use TCP/IP to transmit and receive calls as with DT-24+, DE-30+, and VG200.
- ☑ MGCP is a master/slave protocol, where the gateway is the slave servicing commands from the master, which is the call agent. The MGCP protocol functions in an environment where the call control intelligence have been removed from the gateway.
- ☑ Session Initiation Protocol (SIP) is an application layer control protocol that can establish, modify, and terminate multimedia sessions or calls.

Choosing a Voice Gateway Solution

- ☑ Determining the right voice gateway solutions will depend on a number of factors, from the size and scale of the organization to the budget.
- ☑ Solutions from a switch point-of-view would include, the Catalyst 4000, 4224/4248, and 6000 family. If you wish to use routers, you should choose from the following: the 1750, 2600, 3600, 3810, 7200, and 7500 Series. Access servers may be best in some instances, including the AS5300, the AS5400, and the AS5800. Cisco DT-24, DE-30, and VG-200 would suffice for standalone protocol solutions.
- ☑ For small- to mid-sized companies looking for a nice all-in-one solution, the ICS 7750, deployed with a Catalyst 3524XL-PWR switch and Cisco IP phones, would do wonderfully.
- ☑ The DPA 7610/7630 Voice Mail Gateway would be another important element of an AVVID solution. It provides a gateway allowing legacy voice mail systems to communicate with Cisco CallManagers.

Choosing a Video Gateway Solution

- ☑ Cisco's family of video gateway solutions can satisfy everyone from the small 40-person organization to those with 4000 employees.
- ☑ The IP/VC 3510 MCU connects three or more H.323 videoconference endpoints into a single multiparticipant meeting and is able to support ad-hoc or scheduled videoconferences. Participants can join by having the MCU dial to them or by using the Web interface.
- ☑ IP/VC 3520 and 3525 gateways provide the translation services between H.320 and H.323 networks. This system allows users to conduct videoconferencing across the IP LAN, or via the PSTN. The IP/VC 352x series gateways support V.35, ISDN BRI, and ISDN PRI interfaces. IP/VC 3530 VTA translates from a H.320 ISDN-based system to a H.323 IP-based network. The IP/VC 3540 solution is a highly scalable MCU, which is chassis-based and expands to up to three modules. These modules come in 30-, 60-, and 100-user versions.

Multimedia Conference Manager Services

- ☑ Multimedia Conference Manager (MCM) works in conjunction with Cisco's IP/VC products, and services a H.323 gatekeeper and proxy.
- ☑ MCM is a part of the Cisco IOS for the following router platforms: 2500, 2600, 3600, 3810, and 7200.
- ☑ The MCM gatekeeper functions include: zone administration, RAS, AAA services, bandwidth management, session management, and call accounting. The proxy service provides QoS capabilities to the videoconferencing sessions.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: When should I use a H.323-based gateway versus an MGCP-based gateway?

A: Choose a H.323-based gateway if you are using a Catalyst 4000 AGM module or have a 1750, 2600, or 3600 with an older version of IOS.

Q: Is there an all-in-one gateway, router, and voice-capable switch solution?

A: The Catalyst 4224 switch is able to provide connectivity to the IP WAN and PSTN for gateway and router functionality. It can also provide switched inline power Ethernet ports for IP phones.

Q: What gateway will support QSIG PRI PBX-to-PBX signaling?

A: The Cisco 2600, 3600, MC3810, AS5300, and 7200 Series routers support QSIG PRI, which provides for PBX-to-PBX signaling. This allows enterprises to eliminate expensive lease voice circuits.

Q: What Cisco routers support MCM software?

A: The 25xx, 26xx, 36xx, and 72xx Series routers support the IOS-based MCM software.

Q: How does the Cisco MCM integrate or function with the IP/VC family?

A: The MCM functions as the gatekeeper in a videoconferencing solution.

Chapter 4

AVVID Clustering

Solutions in this chapter:

- CallManager Clustering
- Video Clustering
- Designing Clusters: A Case Study

- ☑ Summary
- ☑ Solutions Fast Track
- ☑ Frequently Asked Questions

Introduction

The AVVID solution provides options that allow scaling and load-distribution for both IP telephony and voice/video conferencing features. *Clustering* is a technique used to enhance both the capabilities of the network as well as the redundancy within the network. With the networking capabilities available today, it is possible to cluster both voice networks and video networks. Clustering techniques allow you to scale your networks and, when the need arises, to add more users or services.

Cisco CallManager clusters can be configured to distribute call processing and device registration between multiple servers on the same segment. Up to eight servers can be part of a cluster, with common database information and real-time device registration data being replicated throughout using intra-cluster communications. With common information shared throughout a number of servers, redundancy is achieved. If a single server fails, another server can transparently take over call processing for a group of devices. For very large or multisite installations, several clusters may be used, with information being shared via inter-cluster communication. However, you should use this solution with care, as many Cisco CallManager features will not function between clusters.

Voice and video conferencing is facilitated by terminals producing voice/video data streams, and by Multipoint Control Units, which control the conference. For voice or video conferences larger than those supported by a single MCU, Cisco offers a feature known as *cascading*. Cascading allows you to cascade two or more MCUs, in order to provide a single larger conference. In addition to providing highly scalable conferences, this also provides load distribution between multiple MCUs, and allows voice and video streams to be localized by the use of MCUs on different segments.

CallManager Clustering

CallManager clustering is a method of seamlessly distributing call processing throughout a converged IP network. By using clustering, several CallManager servers can share the burden of call processing, which becomes particularly important in larger or widely distributed IP Telephony implementations.

A *cluster* is defined as a set of Cisco CallManager servers sharing the same database and resources. The roles described in Table 4.1 can be assigned to members of a CallManager cluster.

Table 4.1 CallManager Roles within a Cluster

CallManager Role	Description
Database publisher server	Makes all configuration changes, and produces call detail records.
TFTP server	Handles downloading of configuration files, ring types, and device operating code.
Application software server	Software installed adds features to the IP telephony solution.
Primary call-processing server	Responsible for call-processing functions.
Backup call-processing server	Responsible for call-processing functions.

Each CallManager in the server may be assigned one or more of these roles, but there is only one database publisher server, and one Trivial File Transfer Protocol (TFTP) server per cluster. You must decide on the level of redundancy and processing distribution required in your installation. For larger installations it is recommended to split the database publisher server and TFTP servers onto different servers.

Why Cluster?

The benefits of implementing CallManager clustering are discussed in depth throughout this chapter. They include the following:

- **Resiliency/Redundancy/Survivability** If a CallManager server in a cluster fails, a different operational server can be used by the client for call-processing functions. The publisher database containing IP device configuration information is replicated throughout the cluster; in the event of a server failure the database is not lost.
- **Scalability** By using several CallManager servers with a common database, call processing and other functions can be distributed throughout the cluster. This will ease the load on individual CallManager servers, and allow for localized processing of calls. Using this feature can help facilitate efficient call processing over large wide area networks (WANs).
- **Feature Transparency** CallManager clustering provides the transparent support of user features across a high-speed campus or metropolitan area networks (MANs).

CallManager Cluster Communications

There are two types of communication between CallManager clusters:

- Intra-cluster communication
- Inter-cluster communication

Intra-cluster communication is defined as communications between CallManager servers in the same cluster. It consists of replication of the CallManager Database as well as dynamic information such as the registration of H.323 devices. Inter-cluster communication deals with communications between different clusters, and is established through the use of inter-cluster trunks.

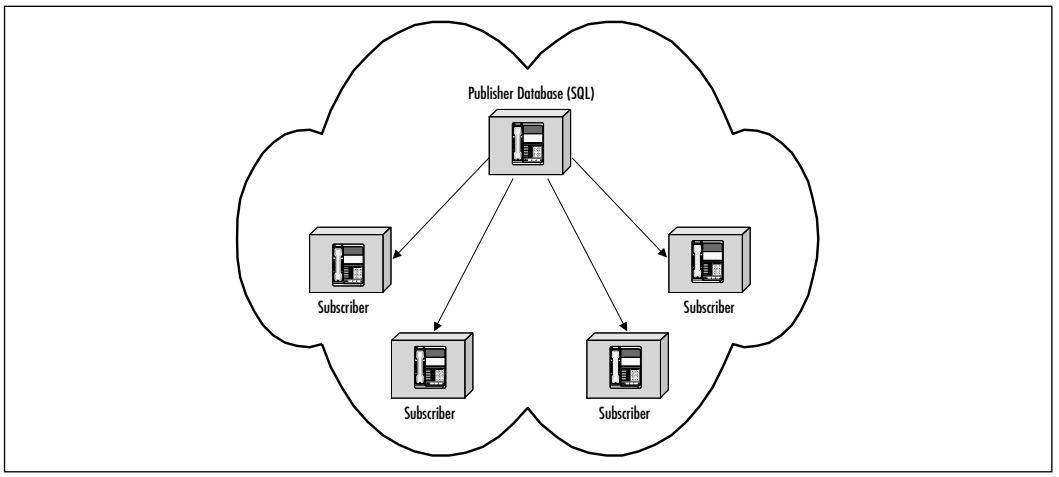
Intra-Cluster Communication

As the name suggests, *intra-cluster communication* is the communication between servers in the same cluster. Two types of information are communicated between servers in the same cluster—CallManager database information, and real-time data.

The CallManager database contains the configuration of all IP telephony devices. When this configuration is updated in the CallManager Administrator, the information is stored in the local database of the CallManager Publisher. The Database Publisher then sends this information to all the database subscribers in the cluster, who then update their local copies of the database.

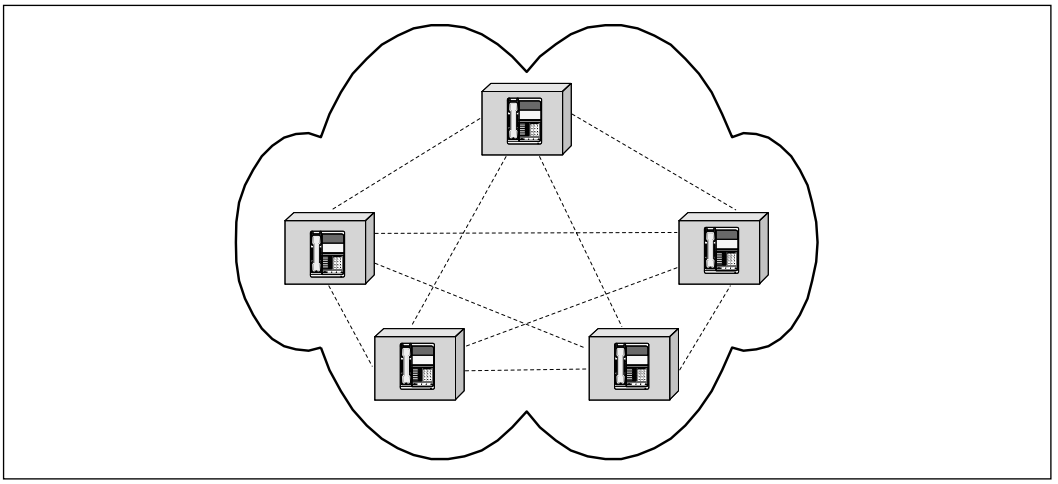
As you can see in Figure 4.1, CallManager database replication works as a client/server model. This feature allows for consistency between databases, and redundancy in the event of a server failure.

Figure 4.1 Intra-Cluster Database Configuration



The real-time data that is communicated by servers within the cluster consists of information such as the registration of IP phones, H.323 gateways, and digital signal processors (DSPs). This information is replicated using a full-mesh, peer-to-peer model (see Figure 4.2) to ensure consistent and optimum routing of calls throughout the system.

Figure 4.2 Intra-Cluster Real-Time Communication



Inter-Cluster Communication

Inter-cluster communication is often required in very large, or widely distributed installations, where more than one CallManager cluster is required to handle the load. H.323 inter-cluster trunks are established to communicate between the different clusters.

The latest release of Cisco CallManager can support up to eight servers within a cluster, and all members of the cluster must reside on the same LAN. However, even though an individual CallManager server can support up to 2,500 IP phones, there is a fixed limit of 10,000 IP phones per cluster. From these limitations, it is obvious that several clusters would be required in very large local networks, or any geographically dispersed network.

There are three types of implementations that might require multiple clusters as well as inter-cluster communication:

- **Large MANs** These are expected to have ample bandwidth to carry the CallManager traffic, and therefore require no call admission control.

- **Multisite WANs with distributed call processing** (At least one CallManager at remote sites) Gatekeepers are used to provide call admission control.
- **Multisite WANs with centralized call processing** (No CallManager at remote sites) The Cisco CallManager locations feature is used to implement call admission control.

Unfortunately, most Cisco CallManager features do not function between CallManager clusters. You must therefore consider that only the following features will exist between clusters:

- Basic call setup
- G.711 and G.729 calls
- Call transfer
- Call park
- Call hold
- Calling line ID
- Multiparty conference

For this reason, you must take care when designing your IP Telephony network, in order to reduce unnecessary CallManager clusters.

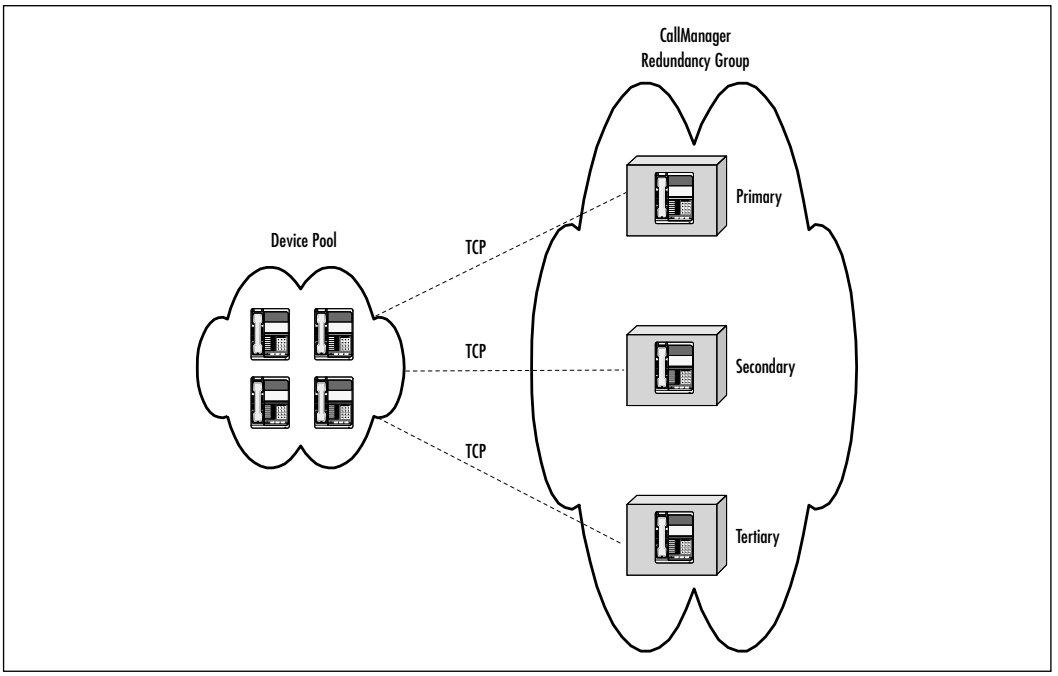
Redundancy within a CallManager Cluster

There are two types of redundancy within a CallManager Cluster: *database redundancy* and *server failover*. Database redundancy is achieved by the replication of the publisher database using the Publisher/Subscriber relationship. This ensures that the same database is held by all servers within a cluster, which means the database is accessible even if a server is not operational. Server failover occurs when a CallManager server being used by an IP Telephony device fails. The device then uses a predefined alternative server to carry out the required tasks. When the preferred server becomes available again, the device switches back to using it.

Redundancy is achieved by configuring CallManager redundancy groups, and then assigning pools of devices to use these groups. A CallManager redundancy group consists of up to three servers—a primary, secondary, and tertiary. If the primary server fails, the secondary server is used by a device in an associated pool, and if the secondary fails, then the tertiary is used. If at any time the primary server becomes operational again, the devices revert back to using that

server. All members of a redundancy group must also be a member of the same cluster. Figure 4.3 illustrates the relationship between devices in a device pool, and CallManager servers within a redundancy group.

Figure 4.3 Cisco CallManager Redundancy



IP telephones, and other IP telephony devices such as gateways, transcoders, conference bridges, and voice mail ports can be assigned to a device pool. A device pool is defined as a group of IP telephony devices that share the same characteristics. These characteristics are defined when the device pool is created, and include name, region, media resource group list, user hold Music On Hold source, auto-answer feature control, and most importantly the associated Cisco CallManager group. As with CallManager groups, each device within a pool should share a similar characteristic, such as geographical location, or logical subnet.

IP phones within a pool maintain a TCP connection with both the primary and secondary CallManager servers in the CallManager group associated with that pool. This facilitates immediate failover should the primary CallManager server fail.

Balanced Call Processing

You can balance calls between the servers in the CallManager cluster through the use of CallManager groups and device pools, as defined in the previous section. By allocating different devices to different pools, and then assigning these pools to different CallManager groups, you have a very flexible method of achieving the load balancing you require.

It is recommended there be some common characteristic shared by each device in a pool, such as physical location, IP subnet, or device type.

In Figure 4.4, we can see two device groups, and two server groups. Within our server group we have three servers; Server A and B are dedicated call processing servers, and Server C is configured as a combined database publisher and TFTP server, but also has call processing capabilities. Server group 1 defines Server A as the primary server, Server B as the secondary server, and Server C as the tertiary server. Server group 2 defines Server B as the primary server, Server A as the secondary server, and Server C as the tertiary server. Device group 1 uses Server group 2, and Device group 2 uses Server group 1. This means that under normal circumstances call processing would be evenly balanced between Server A and B. Server C would only be used for call processing functions if both Server A and B became unavailable.

Designing CallManager Clusters

The powerful features offered by a CallManager cluster depend heavily on a good design methodology. It is very important for you to have a good understanding of the limitations, and Cisco recommendations, of CallManager clustering, in order to design a quality IP telephony infrastructure. With this in place, you will be able to provide a robust, efficient, and feature-rich IP telephony solution.

With Cisco CallManager 3.0(5) we are allowed a maximum of eight servers per cluster, of which a maximum of six may be used for call processing. The remaining two servers should be used as a database publisher server, and TFTP server. These two roles may be combined on the same server, but it is recommended they reside on separate servers for large installations. There can only be one database publisher server, and one TFTP server per cluster.

Cisco recommends the guidelines shown in Table 4.2 when deciding upon the configuration of a CallManager cluster. It is possible to use fewer servers, but redundancy and load balancing will be reduced. The table also indicates the recommended redundancy group configuration. A maximum of 2,500 IP phones, gateways, and DSPs can be allocated to each CallManager with a maximum of 5,000 devices in total.

Figure 4.4 Balanced Call Processing

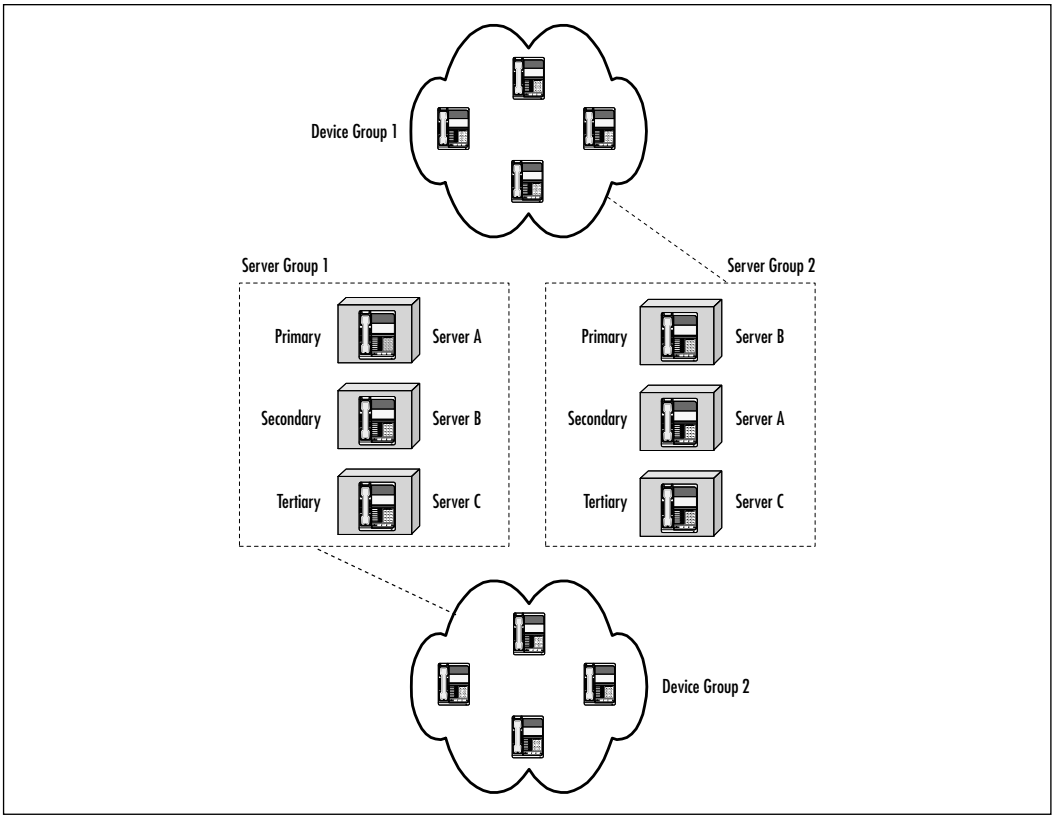


Table 4.2 Recommended CallManager Configurations

IP Phones	Recommended CallManager Configuration	Redundancy Groups
2,500	Three CallManager servers: ■ Database publisher/TFTP server (A) ■ Primary Cisco CallManager (B) ■ Backup Cisco CallManager (C)	One group, servers AB: ■ Server A primary CM for all IP phones ■ Server B backup CN for all IP phones

Continued

Table 4.2 Continued

IP Phones	Recommended CallManager Configuration	Redundancy Groups
5,000	Four CallManager servers: ■ Database publisher/TFTP server (A) ■ Two primary Cisco CallManagers (B and C) ■ Backup Cisco CallManager (D)	Two groups, servers BD, and CD: ■ Server B primary CM for IP phones 0–2500 ■ Server C primary CM for IP phones 2501–5000 ■ Server D secondary CM for all IP phones
10,000	Eight CallManager servers: ■ Database publisher (A) ■ TFTP server (B) ■ Four primary Cisco CallManagers (C, D, E, and F) ■ Two backup Cisco CallManagers (G and H)	Four groups, servers CE, DE, FH, and GH: ■ Server C primary CM for IP phones 0–2500 ■ Server D primary CM for IP phones 2501–5000 ■ Server E backup CM for IP phones 1–5000 ■ Server F primary CM for IP phones 5001–7500 ■ Server G primary CM for IP phones 7501–10000 ■ Server H backup CM for IP phones 5001–10000

Device Weights

Each IP telephony device that registers with CiscoManager is assigned a *weight*. The weight assigned to devices, such as IP phones, H.323 gateways, conferencing hardware, Telephony Application Programming Interfaces (TAPI), and Java TAPI (JTAPI), is based on the memory and CPU resources that they consume. The higher the weight allocated, the more resources they consume.

By considering the weights of devices to be registered in your network, you can work out the number of CallManagers required, and the optimal hardware specification of that hardware. Table 4.3 shows the weights allocated to each device type.

Table 4.3 Weight Chart for IP Telephony Devices

Device Type	Weight per Instance	Instance per Device	Total Device Weight
IP phone	1	1	1
Analog gateway ports	3	Variable	3 per DS0
T1 gateway	3	24	72 per T1
E1 gateway	3	30	90 per E1
Conference resource (hardware)	3	Variable	3 per instance
Conference resource (software)	3	48	144*
Transcoding resource	3	Variable	3 per instance
Software MTP	3	48	144*
CTI port (TAPI or JTAPI)	20	1	20
Messaging (voice mail)	3	Variable	3 per instance
Cisco SoftPhone	20	1	20
Inter-cluster trunk	3	Variable	3

**If installed on the same server as Cisco CallManager, the maximum number of sessions is 48.*

For example, an infrastructure containing 1,250 IP phones, 2 T1 gateways, a software conference resource, and two Cisco SoftPhones, would constitute the following formula: (1,250 x 1) + (2 x 72) + (144 x 1) + (20 x 2) = 1,540 total device weight (device units). Table 4.4 details the maximum number of device units that may be serviced by specific server platforms. Currently, you cannot have more than 2,500 IP phones registered with a single Cisco CallManager, even if the maximum device units allow it.

Table 4.4 Maximum Device Units per Server Platform

Server Platform Specification	Max Device Units per Server	Max IP Phones per Server
MCS-7835-1000 PIII 1000MHz, 1GB RAM	5,000	2,500
MCS-7835 PIII 733MHz, 1GB RAM	5,000	2,500

Continued

Table 4.4 Continued

Server Platform Specification	Max Device Units per Server	Max IP Phones per Server
MCS-7830 PIII 500MHz, 1GB RAM	3,000	1,500
MCS-7830 PIII 500MHz, 512MB RAM	1,000	500
MCS-7825-800 PIII 800MHz, 512MB RAM	1,000	500
MCS-7822 PIII 550MHz, 512MB RAM	1,000	500
MCS-7820 PIII 500MHz, 512MB RAM	1,000	500

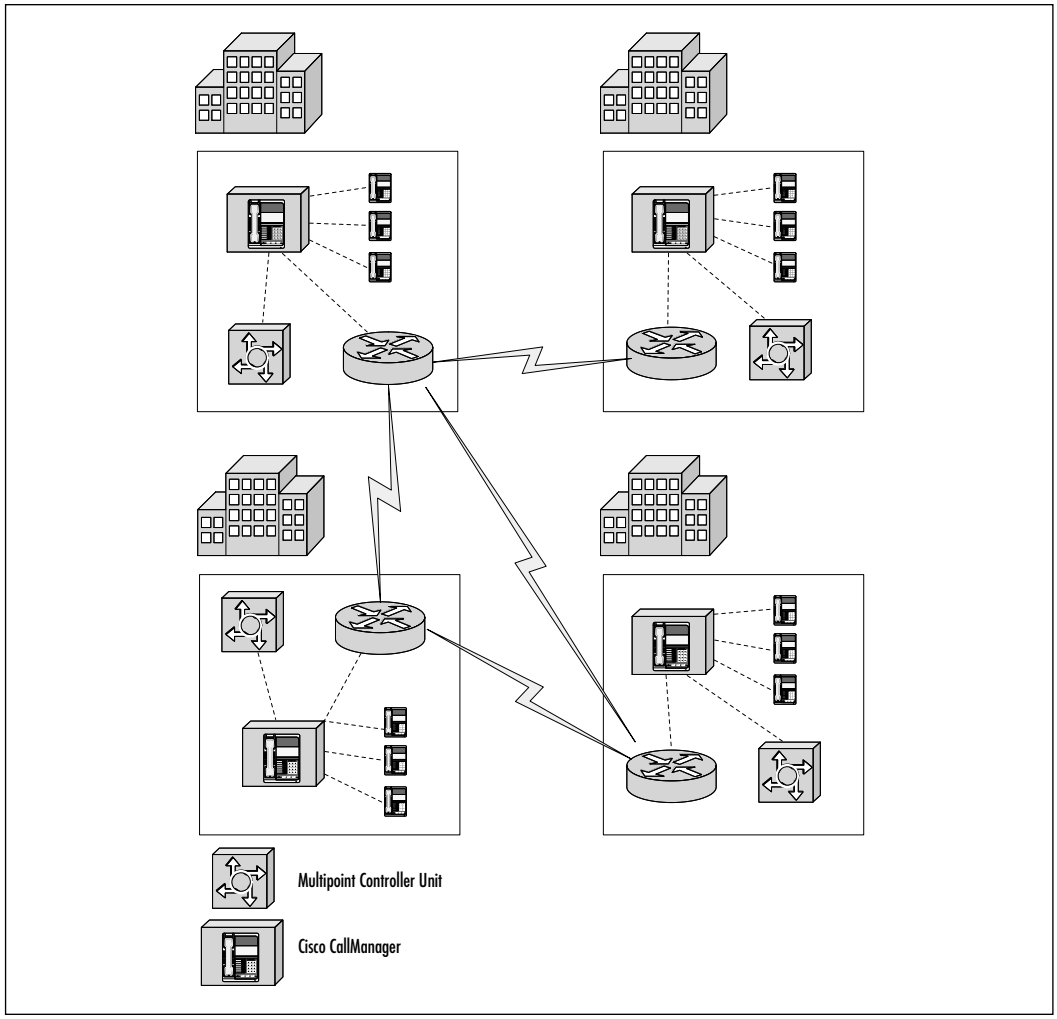
Campus Clustering

As we have discussed previously, it is assumed there is fast, LAN connectivity available between servers within a campus. This is essential because with Cisco CallManager 3.*x*, clustering across a WAN is not supported. For most campus-based networks, a single cluster solution is adequate. Call admission control is not required within a campus, but the following restrictions do apply:

- Maximum of 10,000 total registered devices
- Maximum of eight servers per cluster
- Maximum of 2,500 registered IP phones, or 3,000 other devices per CallManager
- Switched infrastructure to the desktop

You should ensure that the maximum redundancy and load-balancing options are provided by your cluster. It is important to consider that typically, some sites within a campus will have only a single high-speed IP link to the rest of the MAN. In such cases, it is essential that CallManagers are placed on site, to ensure system availability in the event of a link failure. Figure 4.5 illustrates a typical MAN, and how clustering might be used to ensure a high-availability and high-performance system.

Figure 4.5 Clustering over a MAN



Guidelines for Multiple Clusters

Earlier in this chapter we discussed a couple of scenarios where it would be necessary to use multiple clusters. These were for very large networks with over 10,000 registered devices, and for IP Telephony networks distributed over wide area links. Remember that clustering is not supported over a WAN. Communications between clusters require H.323 inter-cluster links.

There are three multicluster designs that may be tailored to fit your design goals:

- Multiple clusters within a campus or Metropolitan Area Network (MAN)
- Multiple clusters over a multisite WAN with distributed call processing
- Multiple clusters over a multisite WAN with centralized call processing

When using multiple clusters on a MAN, the following should be considered:

- Call admission control is not required over a MAN.
- Cisco recommends a maximum of two inter-cluster peers per device.
- Where a gatekeeper is used, Cisco recommends a single H.323 connection per cluster. Implement redundancy by assigning a CallManager redundancy group to the gatekeeper.

Configuring & Installing…

Cluster Configuration Checklist

The following is a checklist to configure a cluster:

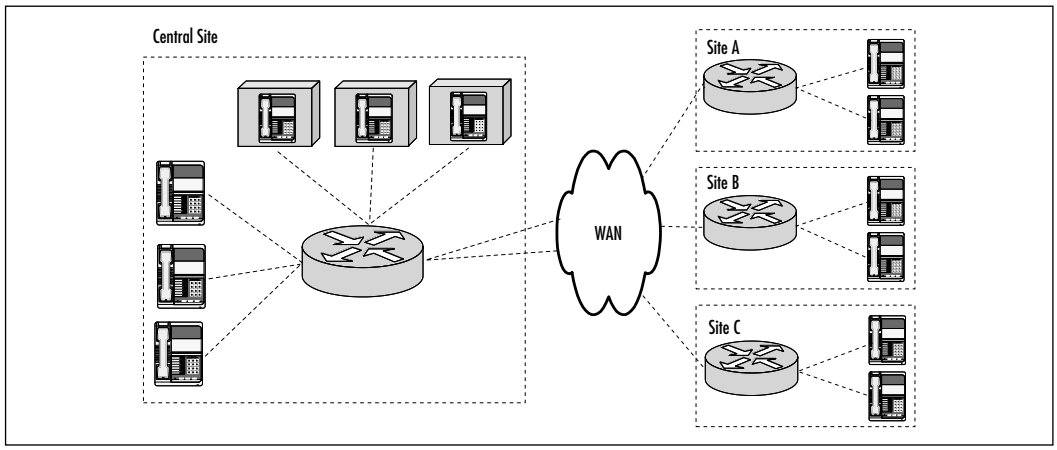
1. Install and configure the servers and other hardware in the cluster.
2. Collate information required to install Cisco CallManager as well as other applications to be installed on servers in the cluster.
3. Decide upon the role of each server within the cluster.
4. Install Cisco CallManager as well as other software required on the servers.
5. Configure CallManager groups to provide server failover and redundancy.
6. Configure device pools, place devices in these pools, and assign to a CallManager group.
7. If an inter-cluster trunk is required, install and configure it as an H.323 device.
8. If call admission control is required for an inter-cluster trunk, configure either a gatekeeper, or the CallManager locations feature.

If an inter-cluster link is established across a WAN, this becomes a bottleneck for communication between clusters, and the WAN link should be scaled to deal with this extra traffic.

For a network with several large sites, call processing should be distributed between the sites, with H.323 Cisco IOS gateways acting as a gatekeeper. The gatekeeper provides both inbound and outbound access control and can control the amount of bandwidth on the WAN link used by voice applications. Call processing is performed within each site, with calls traversing the WAN as needed.

If you have a network with a single central site, and multiple small branch offices or telecommuters, remote call processing might be a suitable option. All call processing will take place at the central site, and each remote site is configured as a Cisco CallManager location (see Figure 4.6). CallManager tracks both available and used bandwidth for each location, and will permit or deny calls based on this information. With Cisco CallManager 3.*x*, up to 2,500 remote devices can be configured. A single CallManager cluster, containing a single CallManager server must be configured at the local site, specifically to handle these remote sites.

Figure 4.6 Call Admission Control Using Locations



Video Clustering

Videoconferencing is an extremely popular technology that is further facilitated by the ITU H.323 standard offering greater flexibility, scalability, and improved cost-effectiveness. Voice and videoconferencing using H.323 is facilitated by MCUs, voice, and, optionally, video-capable terminals, video-aware gateways, and gatekeepers.

Multipoint Controller Units

A Multipoint Controller Unit (MCU) consists of a Multipoint Controller (MC) and zero or more Multipoint Processors (MPs). The MC is the conference controller, and handles all negotiations between terminals, along with conference resources. The MP directly controls the media streams, and performs switching, mixing, and other audio/video processing.

Cisco offers the Cisco IP/VC 3510 MCU as its MCU solution for voice and videoconferencing. This enables conferences between three or more endpoints, and can support up to 15 simultaneous users. Users can spontaneously set up conferences by dialing the MCU, which automatically sets up the conference. Other users can either be added by the initial user, or dial into the conference themselves. For smaller networks, the IP/VC 3510 can also act as a simple gatekeeper, and supports the cascading of MCUs to facilitate larger conferences. Table 4.4 shows the maximum number of conference participants supported by a single IP/VC 3510MCU.

Table 4.4 Maximum Participants per MCU

Participants	Bit Rate	Conference Type
24	64 Kbps	Voice only
15	128 Kbps	Multimedia
9	384 Kbps	Multimedia
5	768 Kbps	Multimedia
3	1.5 Mbps	Multimedia

For larger organizations requiring increased scalability, Cisco offers the second generation Cisco IP/VC 3540 MCU. The 3544 Chassis provides power load balancing and redundancy, status LEDs, and cooling fans, and can accommodate up to four IP/VC 3540 MCU cards on a cPCI bus. The first MCU in the chassis must be a system module, which provides management functions for modules installed in the remaining slots. An application server module is also available, which is a Windows NT system installed with Data Collaboration Server (DCS). The T.120 based IP/VC DCS allows users to present slides, use a whiteboards and graphics, and collaborate with other applications during a conference. Table 4.5 shows the maximum number of conference participants supported by each IP/VC 3540 MCU module, and if they provide system management functions.

Table 4.5 IP/VC 3540 MCU Modules

Module	Participants	Management
IP/VC-3540-MC03S	30	System
IP/VC-3540-MC06S	60	System
IP/VC-3540-MC10S	100	System
IP/VC-3540-MC03	30	Nonsystem
IP/VC-3540-MC06	60	Nonsystem
IP/VC-3540-MC10	100	Nonsystem

Cascading MCUs

As we can see, the limit of 15 conference users per MCU is not adequate for many larger conferences. To overcome this limitation, we can use a feature known as *cascading*. Cascading allows two or more conferences managed by separate MCUs to be joined together, in order to produce a much larger conference. Participants of each conference, meanwhile, will be unaware of the cascaded nature of the conference.

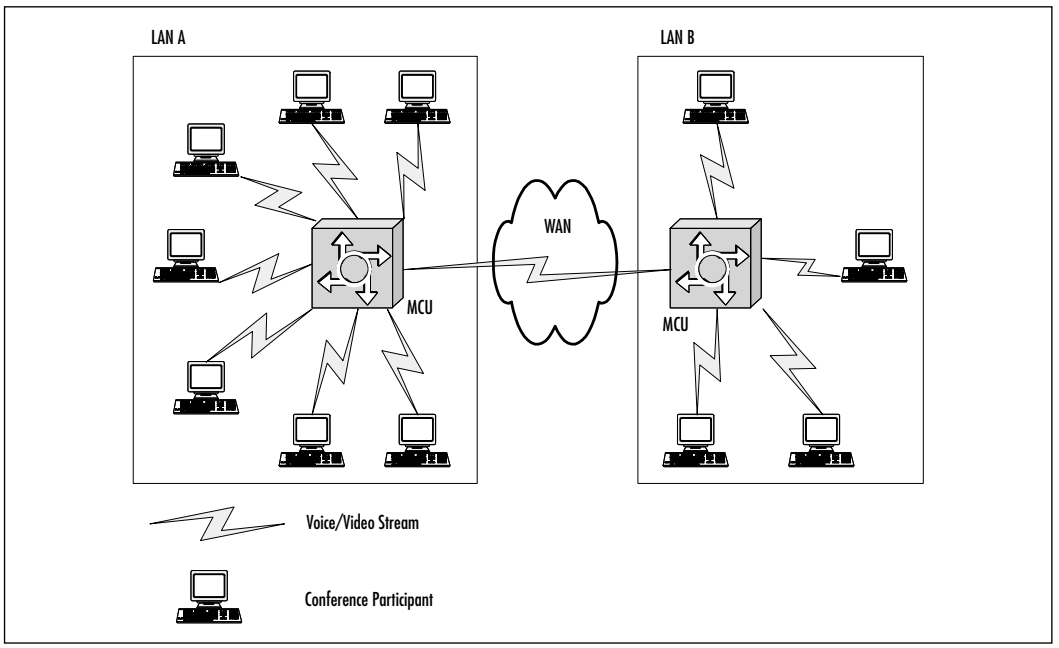
Cascading MCUs is also an excellent way of distributing processing load, and enabling the local processing of voice/video data streams. When a user joins a cascaded conference, she attaches to the nearest MCU. Each individual participant using this particular MCU will have their own voice/video stream to the MCU. However, if there is another remote MCU with another group of participants that is part of the same conference, only a single voice/video data stream flows between the MCUs. A MCU carries out processing functions for only the participants attached to it.

The number of MCUs that can be cascaded into the same conference depends on available bandwidth. By placing MCUs on different segments, it is possible to reduce network congestion caused by voice/video traffic, and to increase the potential number of participants in a conference.

Once a conference is established on a MCU, other MCUs are added by invitation. Once a MCU is invited, and successfully joins a conference, participants using this MCU can invite further MCUs to join. Figure 4.7 illustrates a conference between two cascaded MCUs.

Figure 4.7 shows how bandwidth consumption can be reduced by distributing MCUs across different network segments. In this example, seven participants are connected to the MCU on LAN A at 384 Kbps, with four on LAN B at 384 Kbps. The MCUs are cascaded together, totaling eleven participants at 384 Kbps. This would not be possible if a single MCU were used, for it would only allow a maximum of nine users at 384 Kbps. The options available would be either to reduce the video quality, or, as in this case, to cascade two MCUs. However, the major advantage of this solution is that by using two MCUs we are reducing the communications across the wide area link to a single stream. If a single MCU at LAN B were used, the seven participants on LAN A would have to register with it. This would mean 7 x 384 Kbps = 2,688 Kbps of traffic traversing the wide area link.

Figure 4.7 Cascaded MCUs



Configuring & Implementing…

Setting Up a Cascaded Conference

The following are prerequisites for a cascaded conference:

- Each individual conference must have the same or similar video bit rates and frame rates.
- Prefixes of conferences provided by the MCU must be unique.
- All MCUs participating in the cascaded conference must register with the same gatekeeper or neighbor gatekeeper.

To create a cascaded conference, you must invite a MCU to join a conference managed by another MCU. If a participant joins a conference on the invited MCU, it actually joins the cascaded conference, and can exchange voice and video as usual. There are two methods of inviting a MCU to join a conference:

- By using the MCU monitoring screens through a Web browser.
- By using a terminal to invite an MCU to join when dialing into a conference.

To invite a MCU to join a conference using a terminal such as an IP phone, dial the conference password of the host MCU, the invite string **, then the conference password of the invited MCU.

For example, to invite a MCU whose conference password is 828 to join a conference hosted by a MCU with the password 007, dial 007**828. Other MCUs may then be invited into the conference by terminals connected to the cascaded conference.

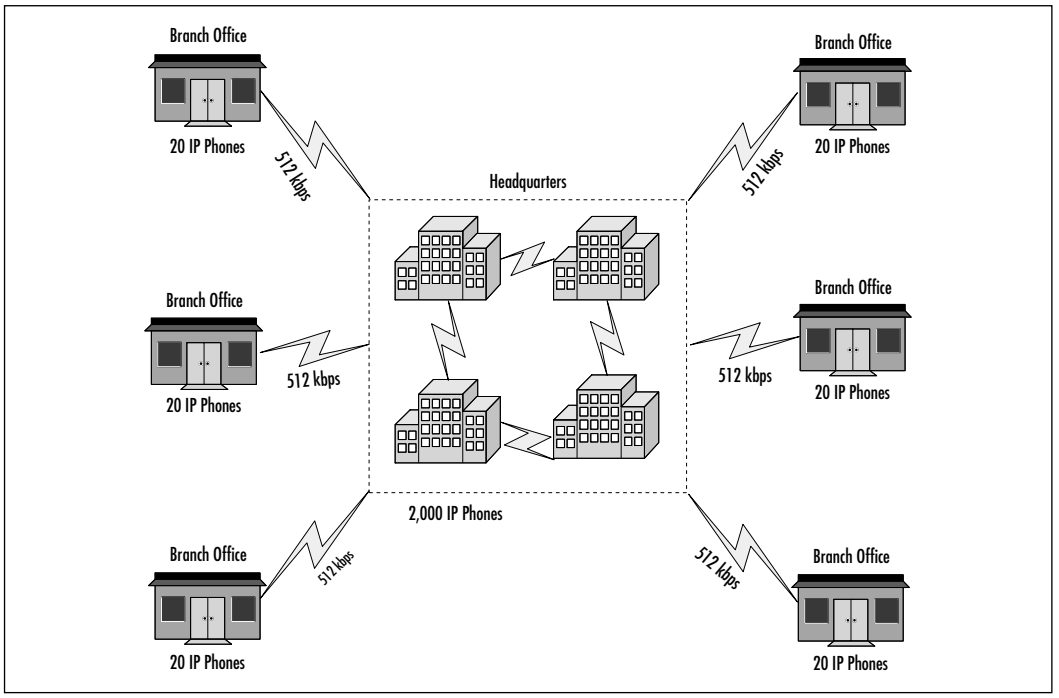
Designing Clusters: A Case Study

In this section, we will consider a hypothetical situation where a customer has come to your team with an IP telephony requirement, along with details of their existing network. Your role in the team is to recommend a robust, scalable, and cost-effective solution.

Gathering Background Information

The customer is a finance house with an enormous headquarters, and 20 small branch offices. The headquarters is located in several buildings on the same campus and is connected by high-speed links. Each branch office is connected via a 512 Kbps Frame Relay circuit. Figure 4.8 illustrates the original network design.

Figure 4.8 Case Study: Corporate Network



The IP telephony solution suggested by the team includes the following requirements:

- 2,000 IP phones at the main headquarters.
- 40 IP phones at each of the 20 branch offices.
- Half of the IP phones will have voice mail accounts
- Provision for 50 analog gateways
- Two T1 gateways
- Videoconferencing facilities for all users. It must be possible for users to conference between branch offices, and some conferences can involve up to 30 participants.

The customer does not want to spend more than what's necessary, but does not care to cut costs at the expense of performance or functionality.

Coming to a Possible Solution

This section will detail one possible solution to the previous requirement. You must remember that there are potentially several solutions that will meet all the requirements, and you should involve the customer when deciding which is the most suitable option for them. Of course, each possible solution will have associated benefits, restrictions, and costs.

What Are the Videoconferencing Requirements?

Let's first consider the videoconferencing requirement. The design team has already given you specific requirements for the IP telephony devices, but it is up to you to decide the optimum conferencing solution. The main requirements are to be able to support up to 30 users per conference, and support conferences between sites.

A single MCU can support 15 multimedia participants at 128 Kbps; therefore, to support 30 participants you require at least two. It would be possible to fulfill their requirements by placing two MCUs at the headquarters. However, as each conference participant requires a 128 Kbps data stream, the 512 Kbps link to remote sites would quickly become saturated. Also, if a conference was being held within a remote site, all conference traffic would have to cross the WAN link to the headquarters and back again. Therefore, you would recommend placing a MCU at each remote site in addition to the two at the headquarters. This would ensure that all conference traffic within an office is kept local, and only one data stream exists between MCUs, saving valuable WAN bandwidth.

Does the Customer Need Clustering?

Your next task is to decide whether or not clustering is required, and what should be the configuration of these clusters. The first step in this process is to total up the number of IP phones required, as well as the total device weight of IP telephony devices on the network.

There are 2,000 IP phones on the headquarters campus, and 40 IP phones at each of the 20 remote sites. Using the formula, that tabulates as 2,000 + (20 x 40) = 2,800 IP phones.

Device weights are as follows:

- 2,800 phones @ 1 unit each = 2,800 units
- 1,400 voice mail accounts @ 3 units each = 4,200 units
- 50 analog gateways @ 3 units each = 150 units
- 2 T1 gateways @ 72 units each = 144 units
- 22 MCUs @ 45 units each = 990 units

This gives a grand total of 8,284 device units.

A single CallManager server will support up to 2,500 IP phones, and you require 2,800, so you can immediately see that you need to use clustering. Even if this were not the case, a single CallManager server can only support 5,000 device units, and you require 8,284. In addition, you require clustering to facilitate the customer's reliability requirements.

Does the Customer Need Multiple Clusters?

You now need to decide how many clusters you need. A single cluster will support a maximum of 10,000 IP phones, so your total of 2,800 falls well under this. However, a cluster is only supported across fast LAN media and you have 20 remote sites supported across slow frame relay links. You have two options, the first being to use multiple clusters, with one at the headquarters, and one at each remote site. This could be seen as wasteful since each remote site would have at least one CallManager server supporting only 40 clients. In addition, you would lose a lot of CallManager functionality, as many CallManager features are not supported between clusters.

A preferable option would be to use a single cluster located at the central site, with each remote site defined as a location within CallManager. This would allow CallManager to track and control voice traffic to remote sites without an expensive and complex multicluster solution.

Cisco also offers a new IOS feature called Survivable Remote Site (SRS) telephony. This feature allows a Cisco router at a remote site to automatically detect a failure, and provide call processing functionality to IP phones for the duration of the failure. When the link is restored, call processing automatically switches back to Cisco CallManager. This ensures that local calls at small sites without Cisco CallManager can still be processed in the event of a failure in the link to the central site. Configuration is performed on the Cisco CallManager, requiring little or no administration at the remote site. SRS is available on Cisco 2600 and 3600 routers, and Cisco Catalyst 4224 Access Gateway Switches.

What Hardware Is Required?

Since you require 2,800 IP phones and 8,284 device units, under Cisco recommendations you would need four CallManager servers which could support up to 5,000 phones. This combination would allow for further growth, as well as improved performance and reliability.

You should suggest using four Cisco MCS-7835-1000, PIII 1000MHz, with 1 GB RAM. This would allow for a maximum of 10,000 IP phones, and 20,000 device units. If cost were a primary issue, lower specification servers could be used instead to achieve your design goals.

How Is Redundancy Achieved?

If you follow Cisco recommendations, your cluster would be configured as follows:

- One Database publisher/TFTP server (A)
- Two primary Cisco CallManagers (B and C)
- One backup Cisco CallManager (D)

You would require two CallManager groups for redundancy (configured as follows):

Group	Primary CallManager	Secondary CallManager
BD	B	D
CD	C	D

You then need to configure half the IP phones to use group BD, and half to use CD to facilitate load balancing. You would achieve this by configuring the following device groups:

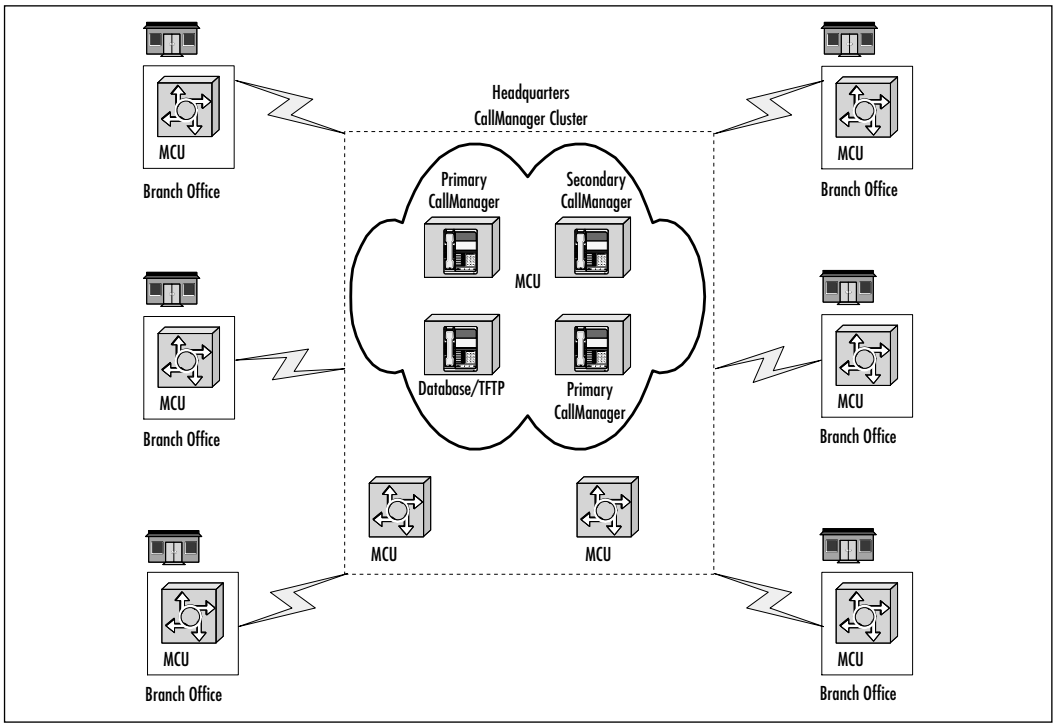
Device Group	IP Phones
A	1 to 1,400
B	1,401 to 2,800

Configuration Summary

Although there are several possible designs that achieve the required goals, this would be a very effective solution. Reliability and load balancing would be achieved by configuring two CallManager groups using different primary servers in each as well as a shared backup server, and assigning half the IP phones to each. Scalability is provided by spare capacity, and the ability to add servers to the cluster. Using CallManager locations for remote sites would allow for tracking and control of traffic over the WAN, and would be much less complex than using several clusters.

Videoconferencing facilities are provided through the installation of MCUs. A MCU at each remote site means that all intra-branch conference traffic will be kept local. If conferences take place between sites, only the MCUs communicate with each other, thus saving WAN bandwidth. The provision of two MCUs at the central site will allow for up to 30 users within that site alone. See Figure 4.9 for the solution realization.

Figure 4.9 Case Study: Design Solution



Summary

Throughout this chapter we have discussed the importance of redundancy and load distribution in an AVVID infrastructure, and how this may be achieved using features supported by Cisco equipment.

We have looked at how several Cisco CallManagers can be *clustered* together on a fast network to provide transparent redundancy, and illustrated load balancing of call processing between each server.

Intra-cluster communication takes place between members of the same cluster. Information communicated using this method includes the CallManager Database, replicated using a client server model, and real-time data such as device registration, using a full-mesh topology. Intra-cluster communication requires a fast, switched network media to perform.

There is a limitation of eight CallManager servers in a cluster, which must include a database server, TFTP server, a primary call processing server, and backup call processing servers (optional), and application servers. If necessary, these features can be combined onto the same server. The number of devices supported by each server depends on the processor, the memory of the server, and the weights of the devices connecting. Regardless of the hardware, no more than 2,500 IP phones can be supported by a single CallManager server, and no more than 10,000 IP phones can be supported by a single cluster.

To overcome these limitations, it is often necessary to use multiple clusters within an organization if it is very large or geographically diverse. Inter-cluster communications take place between separate clusters using the H.323 protocol over inter-cluster trunks. Unfortunately, many CallManager features are not supported between clusters. The only features currently supported are basic call setup, G.711 and G.729 calls, call transfer, call park, call hold, calling line ID, and multiparty conference. Therefore great care should be taken when designing the IP telephony network, and unnecessary clusters should be avoided.

The use of Cisco CallManager groups facilitates redundancy, and load balancing between cluster members can be achieved using groups and device pools. Three CallManager servers—a primary, secondary, and tertiary—can be part of a group. A device pool is a group of IP telephony devices sharing the same characteristics, including the CallManager group associated with it. When a device or device pool is associated with a CallManager group, it uses the primary server when available, but can switch to using the secondary or tertiary server in the event of a failure. Load balancing can be achieved by assigning different device pools to different CallManager groups.

Voice and videoconferencing devices may also be linked together to improve scalability and performance through a feature known as *cascading*. Two or more Media Control Units can be linked together in order to facilitate larger conferences and to effectively distribute conference processing and voice/video traffic. Users only share voice/video streams with the MCU to which they are attached, and there is only a single voice/video stream between each MCU. Conference participants can easily invite other MCUs to become part of the conference through the use of their terminal or by way of a Web browser.

Through the use of features supported by the Cisco AVVID infrastructure, it is possible to implement robust, high-performance, and scalable multimedia solutions.

Solutions Fast Track

CallManager Clustering

- ☑ Cisco AVVID infrastructure includes a variety of features to facilitate load balancing, scalability, and redundancy for IP telephony and multimedia conference solutions.
- ☑ Cisco CallManager clusters are used to improve the scalability and reliability of Cisco IP telephony solutions.
- ☑ Multipoint Control Unit cascading is used to improve the scalability of voice/video conferencing.
- ☑ A maximum of eight Cisco CallManagers can be members of a cluster, with as many as six used for call processing.
- ☑ The possible roles of servers within a cluster are: database publisher server, TFTP server, application server, primary call-processing server, and backup call-processing server.
- ☑ Intra-cluster communications rely on high-speed network connections, and are not supported across WANs.
- ☑ The CallManager database contains the configuration of all IP telephony devices.
- ☑ Real-time data replicated between servers in a cluster consists of registration information of IP telephony devices.

- ☑ Many CallManager features do not function between different clusters.
- ☑ Database redundancy is achieved by replicating the publisher database to all servers within a cluster.
- ☑ Redundancy groups facilitate server failover. A device is associated with a redundancy group, which is a list of up to three servers. If the primary server fails, call processing is transferred to the secondary server.
- ☑ Balanced call processing can be achieved by assigning different primary servers to different groups of devices.
- ☑ Device weights are used to calculate the maximum number of devices that can be supported by a single CallManager server.

Video Clustering

- ☑ A maximum of 15 conference participants can be supported by a single MCU.
- ☑ Two or more MCUs can be cascaded to support larger conferences.
- ☑ Conference participants are unaware of the cascaded nature of the conference.
- ☑ Only a single voice/video data stream exists between cascaded MCUs.
- ☑ Voice/video traffic can be localized by correctly dispersing MCUs across a network.
- ☑ The number of MCUs that can be cascaded together depends on available bandwidth.
- ☑ To invite a MCU to join a conference from a terminal, dial the host conference password, the invite code **, followed by the conference password of the invited MCU.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: Why would I want to consider using Cisco CallManager clustering?

A: CallManager clustering improves both the scalability and reliability of Cisco IP telephony solutions. A maximum of 2,500 IP phones can be supported by a single CallManager server—any more than this and you will require a cluster. Call processing can be distributed throughout the cluster providing balanced call processing. Reliability is increased by the automatic replication of the CallManager database, and redundancy groups can be configured to provide devices with alternative sources of call processing should a server fail.

Q: How do I decide how many servers I should have in my cluster?

A: Cisco offers comprehensive guidelines on how to provision clusters, all of which can be found in this chapter. All clusters require a TFTP server, as well as a database publisher server. On all but the largest networks these can be combined onto the same CallManager server. You will then require one or more primary call processing servers, and optionally a number of backup call processing servers. Each IP telephony device has an associated device weight, these should be totaled together to give the total device weight units of devices on your network. The number of device units supported by a CallManager server depends on the model. You must then use the design guidelines to determine the required level of redundancy on your network.

Q: I would like to configure a CallManager cluster across several remote sites. What should I take into consideration?

A: CallManager 3.0(5) does not support a single cluster across a WAN. You will need to set up multiple clusters if the sites are large, or consider using the CallManager locations feature if the sites are small.

Q: I need to use a several different CallManager clusters across my network. How do they communicate, and are there any limitations I should be aware of?

A: Inter-cluster communication takes place over inter-cluster trunks using the H.323 protocol. Most CallManager features do not scale between clusters; only the following features will be available: basic call setup, G.711 and G.729 calls, call hold, call park, call transfer, calling line ID, and multiparty conference.

Q: I need to make provisions for a conference for nine users split across three sites using 768 Kbps multimedia data streams. How can I find out how many MCUs I require, and where they should be placed?

A: A minimum of two MCUs will be required to support this, as a single MCU supports five participants at 768 Kbps. However, it is worth considering using three, one at each site, to reduce the multimedia traffic over the wide area links.

Q: How do I invite another MCU to join our conference? Our conference password is 263, and the remote conference password is 899.

A: You could either use a Web browser to access the MCU monitoring tools, or the easier option would be to use your IP phone. Dial the host conference password, the invite string **, then the invited conference password. In your case, this would be 263**899.

Chapter 5

Voice and Video Gatekeeper Design

Solutions in this chapter:

- Understanding Gatekeeper Basics
- A Gatekeeper's Role in Voice and Video Networking
- Placing and Configuring Gatekeepers: A Case Study

- ☑ Summary
- ☑ Solutions Fast Track
- ☑ Frequently Asked Questions

Introduction

Gatekeepers are an essential component when designing AVVID networks. From a videoconferencing perspective, the gatekeeper is the device that will permit or deny requests for videoconferences, making the judgment as to whether there are enough resources to make or accept a specific videoconference connection.

When looking at the Internet Protocol (IP) telephony component of AVVID, gatekeepers are commonly used in multisite distributed call processing scenarios (discussed in greater detail in Chapter 11). As in the videoconferencing design, the gatekeeper has a set of values by which it will determine whether or not to allow a specific call, regardless of whether the call is incoming or outgoing.

By the end of the chapter, you will have an understanding of the gatekeeper's specific functions, where and to place the gatekeepers in an AVVID network, and why, as well as a comprehension of design considerations when placing gatekeepers for videoconferencing or for IP telephony purposes.

Understanding Gatekeeper Basics

This section of the chapter discusses the functionality of the gatekeeper and the purposes the gatekeeper serves. It examines the types of gatekeepers available and the way the gatekeeper interacts with other devices on the network. It also covers design considerations, examining options that should be considered when designing your voice and video network.

What Is a Gatekeeper?

The gatekeeper acts as an intelligent, central point of control for a real-time, multimedia (H.323) network. It monitors endpoints and gateways as well as audio, video and collaborative data calls. The gatekeeper can control (based on its configuration) what stations (endpoints) participate in the network. It can also restrict calls based on the endpoint that places or receives the call, the time of day, and so on. In addition, it can perform various management functions such as address resolution, directory services, as well as call authorization and accounting.

In most Cisco networks, the gatekeeper is also known as the Multimedia Conference Manager (MCM). This is an IOS-based gatekeeper that runs on many router platforms. The gatekeeper can be configured on an existing router or on a new, dedicated router. Cisco recommends the 2600, 3600, or 7200 platforms for the MCM gatekeeper. As with any function, performance will vary

depending on the platform. (Table 5.3 later in this chapter compares the performance of these three product families.)

Cisco has recently introduced an extension to the MCM, called the High Performance Gatekeeper. This product greatly enhances the scalability and redundancy of the MCM. The final type of Cisco gatekeeper comes with the Cisco Video over IP conferencing (IP/VC) video products and is known as an embedded gatekeeper. (Table 5.1 later in the Gatekeeper Basics section compares the features of the three different types of gatekeepers.)

Gatekeeper Functions

Gatekeepers are a component of an H.323 network—a network designed to transport real-time traffic, such as voice, video, or collaborative data. A gatekeeper interacts with *endpoints*, which are stations capable of placing H.323 calls, such as a workstation running Microsoft NetMeeting or a Cisco CallManager. A gatekeeper also interacts with gateways, which are devices capable of translating H.323 traffic into other forms of traffic, and which were discussed in Chapter 3. For example, gateways convert H.323 traffic into voice calls over the traditional phone network or Integrated Services Digital Network (ISDN) calls, common with videoconferencing. This chapter explores what a gatekeeper is and what functionality it provides.

As defined by the H.323 protocol, the gatekeeper is required to perform a certain set of functions. These required functions perform basic H.323 services. For example, the gatekeeper locates endpoints that are receiving calls, relieving endpoints of this task. The gatekeeper also controls overall participation in the network as well as calls placed there. Additional functions are optional and may add value in certain cases. The next two sections review both types of functions.

Gatekeepers use the H.225 protocol to communicate with endpoints and gateways. The H.225 protocol has two basic parts: Registration, Admission, and Status (RAS) and call signaling. Gatekeepers primarily use the RAS portion of the H.225 protocol with endpoints and gateways for registration, admission, and call control in the H.323 network. Endpoints and gateways also use the call signaling portion of the protocol for call setup and tear down.

Required Functions

Gatekeepers are required to perform all of the following functions. Since endpoints are required to use a gatekeeper if one is available, this is an excellent control point for the network:

- **Address translation** Also known as address resolution, the gatekeeper will translate an H.323 address (such as an E.164 phone number) into an IP address. The gatekeeper will do this by resolving the phone number to an endpoint already registered with the gatekeeper or by finding the location of the phone number by querying other configured gatekeepers using the H.225 (RAS) protocol. For example, the gatekeeper can translate 212-555-1212 into 10.15.6.1. The gatekeeper can also translate based upon H.323 IDs (character strings).
- **Admission control** The gatekeeper can control what endpoints join and participate in the H.323 network. For simplicity, the gatekeeper can be configured to allow all endpoints to join the H.323 network. Alternatively for tighter security it can only admit a known list of endpoints. The gatekeeper may also restrict endpoint participation by other settings configured by the administrator, such as available bandwidth or number of active endpoints. Although an H.323 network does not require a gatekeeper, if a gatekeeper exists, all participants are required to use it, allowing security to be enforced.
- **Bandwidth control** The gatekeeper is responsible for monitoring and controlling the network bandwidth being used by all calls. You can restrict the amount of bandwidth used by voice and video (H.323) calls. This is very important because if more calls are placed than the network can support, all calls will suffer from poor quality. For example, the gatekeeper actively monitors all calls, the bandwidth used by each call (bandwidth requested at setup) and the call signaling between endpoints. The gatekeeper uses this information to prevent the total bandwidth used by voice and video calls to exceed the configured limit for a zone. This assures that all allowed calls receive sufficient bandwidth. Thus the gatekeeper can reject calls if a threshold for H.323 traffic has already been met. In a traditional voice network the channels available on the wide area network (WAN) would limit the number of calls that could be placed. In an IP network, this limit does not exist—thus the gatekeeper must apply this limit.
- **Zone management** Zones are a logical group of devices participating in the H.323 network. The gatekeeper controls the zone—what devices may join the zone, what devices may place and receive calls to or from the zone. As the administrator, you control the number and operation of all zones. It is very easy to control the total bandwidth used by H.323

calls into or out of a zone. This often dictates how zones are created in a network.

Optional Functions

A gatekeeper can implement the following functions. All of these functions, except the supplementary services and directory services, are available with Cisco's Multimedia Conference Manager gatekeeper. The IP/VC embedded gatekeepers do offer call and bandwidth management as well as call forwarding (a supplementary service), though they do not offer authentication, authorization, or directory services. They do provide some call accounting, though only through special third-party software.

You may decide to implement some or all of these functions based on the exact needs of your network. Some functions, such as authorization and accounting, you may not implement initially, but may find useful at a later time.

- **Call control signaling (call routing)** The gatekeeper assists H.323 endpoints and gateways completing calls. It can either operate in direct mode or routed mode. In direct mode the gatekeeper facilitates call signaling directly between the endpoints. In routed mode the gatekeeper receives all call-signaling messages and routes the call signals between itself and each endpoint.
- **Call authorization and authentication** When an endpoint attempts to make a call, it will place the request with the gatekeeper. The gatekeeper can authenticate the endpoint (user) with Terminal Access Controller Access Control System Plus (TACACS+) or Remote Dial-In User Service (RADIUS). The gatekeeper can authorize or reject the call based on the user ID alone or in conjunction with parameters such as time of day, the number being called, and so on.
- **Call management** The gatekeeper maintains information about all active calls. This allows it to perform functions such as knowing when an endpoint is busy and rerouting calls to achieve load balancing.
- **Bandwidth management** The gatekeeper uses bandwidth control to only allow calls for which sufficient bandwidth exists. Optionally, the gatekeeper can limit the bandwidth used by a call to less than was requested at setup. Also, the gatekeeper can work with existing Quality of Service (QoS) mechanisms and servers to achieve optimal performance with H.323 calls.

- **Call accounting** The gatekeeper can maintain records about calls placed. Information such as calling and called endpoint, length of call, and time and date of call can be recorded, which is valuable for security, capacity planning, and budgeting reasons. This function is most easily implemented in conjunction with a TACACS+ or RADIUS server.
- **Directory services** Gatekeepers can maintain or reference databases to assist H.323 users finding one another. They can use databases such as the Internet locator service or the Lightweight Directory Access Protocol (LDAP) to determine a user's phone number.
- **Supplementary services** The H.450 standard specifies call functions commonly found in voice networks. Examples of such functions are call forwarding, call transfer, call hold, call waiting, and so on. Some gatekeepers implement these functions for the endpoints that they serve. For example, your H.323 endpoint receiving voice calls may need to forward calls to your cell phone while you are at another facility. Cisco typically implements these features in H.323 gateways or in CallManager. However, as of 12.1(5)XM the MCM gatekeeper will support a gateway that performs call forwarding or call transfers.

Types of Gatekeepers

As with voice networks, there are several implementations of gatekeepers, both from Cisco and other companies. Cisco employs three types of gatekeepers in H.323 networks: *Embedded gatekeepers*, *MCM*, and a new *high performance gatekeeper*. As discussed earlier, while any standards-compliant gatekeeper should function correctly, Cisco gatekeepers offer several advantages. They have been tested in AVVID implementations, and offer features beyond those defined by the standard. Cisco also offers excellent support.

Several vendors have created gatekeeper implementations that run on Intel and Sun platforms. While these implementations do perform the gatekeeper functions, we highly recommend using Cisco's gatekeeper implementation. This not only assures compatibility with other AVVID components, but also provides additional features.

Multimedia Conference Manager

Cisco's Multimedia Conference Manager can be a gatekeeper for any type of H.323 endpoint. Thus endpoints with desktop videoconferencing systems or local

area network (LAN)-attached video systems for conference rooms or auditoriums can register with MCM just as well as Cisco's CallManager or an IP telephone.

Cisco implements the MCM gatekeeper using the H.323/MCM feature set of its router IOS. The gatekeeper can run on the 2500, 2600, 3600, or 7200 router platforms.

The MCM combines the gatekeeper and proxy services into one product. Although the proxy is a separate function from the gatekeeper, it is worth mentioning since it is included with the MCM. The proxy serves several purposes, but the two most common are security and QoS.

The proxy can provide security by hiding the address of endpoints it serves. Calls are made to the proxy, and then the proxy makes a corresponding call into the endpoint. This is similar to the way a Hypertext Transfer Protocol (HTTP) proxy makes a separate request on behalf of a client.

The proxy can assist with implementing QoS. Since all calls coming from the proxy will originate with the proxy's IP address, it is easier to implement priority queues based on that address. Often, proxies have special QoS features, such as the ability to signal RSVP for its calls.

High-Performance Gatekeeper

In IOS release 12.2(2)T, Cisco introduced a substantial enhancement to the MCM gatekeeper. This new implementation introduces clustering of multiple gatekeepers. This provides greatly improved, carrier class reliability, security, and performance. The high performance gatekeeper is supported on the 2600, 3600, M3810, and 7200 platforms.

Gatekeeper clustering is a Cisco feature that groups multiple gatekeepers logically together. Although only one gatekeeper manages a zone, each gatekeeper shares all its local zone information with the cluster. This allows the cluster to effectively manage each zone. Another feature to increase performance is gatekeeper load balancing. One gatekeeper can dynamically move registered H.323 endpoints to another gatekeeper based on a threshold on the gatekeeper being met. Thresholds can be set on the number of calls, CPU utilization, or memory utilization. This increases gatekeeper scalability as well.

The High Performance Gatekeeper offers performance and reliability increases that appeal to enterprises, though this product also has features targeted to a service provider network. One of these features is a robust, open application programming interface (API). This is designed to allow service providers to develop enhanced voice and virtual private network (VPN) solutions to offer to

customers. Another feature is very detailed call information that can be reported to a RADIUS server for billing purposes.

Embedded Gatekeepers

Some of Cisco's videoconferencing systems, the IP/VC products, come with embedded, or built-in gatekeepers. These are ideal for small networks, and perform all of the required gatekeeper functionality (address translation, admissions control, and bandwidth control).

The embedded gatekeeper is compatible with the MCM gatekeeper. Thus, for larger networks, you can have the embedded gatekeeper interoperate with MCM, or simply have the IP/VC products register directly with one of your MCM gatekeepers.

Comparing Cisco Gatekeepers

The Cisco MCM, IP/VC embedded, and High Performance gatekeeper all offer different features. Table 5.1 compares many different attributes across each of these three platforms.

Table 5.1 Comparison of Cisco Gatekeepers

Feature	IP/VC Embedded Gatekeeper	MCM Gatekeeper	High Performance Gatekeeper
Performance	Good	Very Good	Excellent
Target Network Size	Small to Medium	Large	Very Large
Supports "Required" Gatekeeper Functionality	Yes	Yes	Yes
Supports Bandwidth Limits by Zone	Yes	Yes	Yes
Intended for Voice and Video Support	No	Yes	Yes
Supports Authentication and Authorization	No	Yes	Yes
Supports Gatekeeper Clustering	No	No	Yes
Supports Dynamic Load Balancing	No	No	Yes

Continued

Table 5.1 Continued

Feature	IP/VC Embedded Gatekeeper	MCM Gatekeeper	High Performance Gatekeeper
Support for Enhanced API	No	No	Yes
Call Accounting Information Available	Requires a Third-Party Software Product	Moderate Information Available	Detailed Information Available

As Table 5.1 implies, the IP/VC Gatekeeper was intended for small to medium size video networks. Although they can service voice calls, the MCM Gatekeeper is much better suited for that purpose.

The MCM Gatekeeper is sufficient for many enterprise networks where H.323 is just being introduced or is not yet mission-critical. For service providers or organizations where critical voice and video calls are being placed, the High Performance Gatekeeper configured in a cluster is the best solution.

Gatekeeper Flow Diagrams

The RAS portion of the H.225 protocol is defined by requests and responses that follow similar formats. Requests always end in the letters "RQ" which indicate *request*. Responses always end in "CF" which indicate *confirmation*, or "RJ" which indicate *rejection*. The letter or letters preceding these indicate the actual subject. Thus "RRQ" indicates *registration request*, "LCF" indicates *location confirmation*, and so on.

The process of gatekeeper discovery, registration, and call signaling for IP phones using CallManager is shown in Figure 5.1.

Both endpoints (in this case, CallManagers) discover and register with their gatekeeper (Steps 1 to 4). When Phone 1 places a call, its CallManager sends an admission request to its gatekeeper to determine if it may place the call (Steps 5 to 6). The CallManager will usually send a bandwidth request to specify the bandwidth required for the call. Gatekeeper 1 uses a location request to locate the endpoint for the call (Steps 7 to 8). The CallManager receiving the call (on behalf of Phone 2) sends an admission request to its gatekeeper to determine if it may receive the call (Steps 9 to 10). Once the placing and receiving of the call has been approved, actual call setup does not involve the gatekeepers (Steps 11 to 12).

As discussed earlier, the gatekeeper can reject a call based on many factors, such as available bandwidth. Figure 5.2, Gatekeeper Call Rejection, displays an example of this.

Figure 5.1 RAS Signals in an H.323 Voice Network

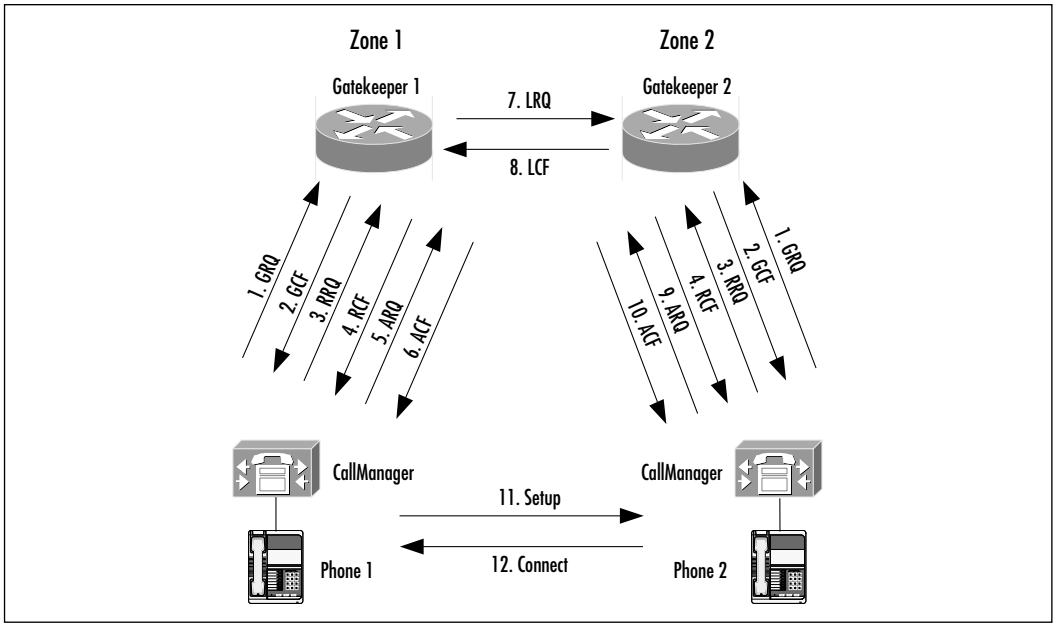
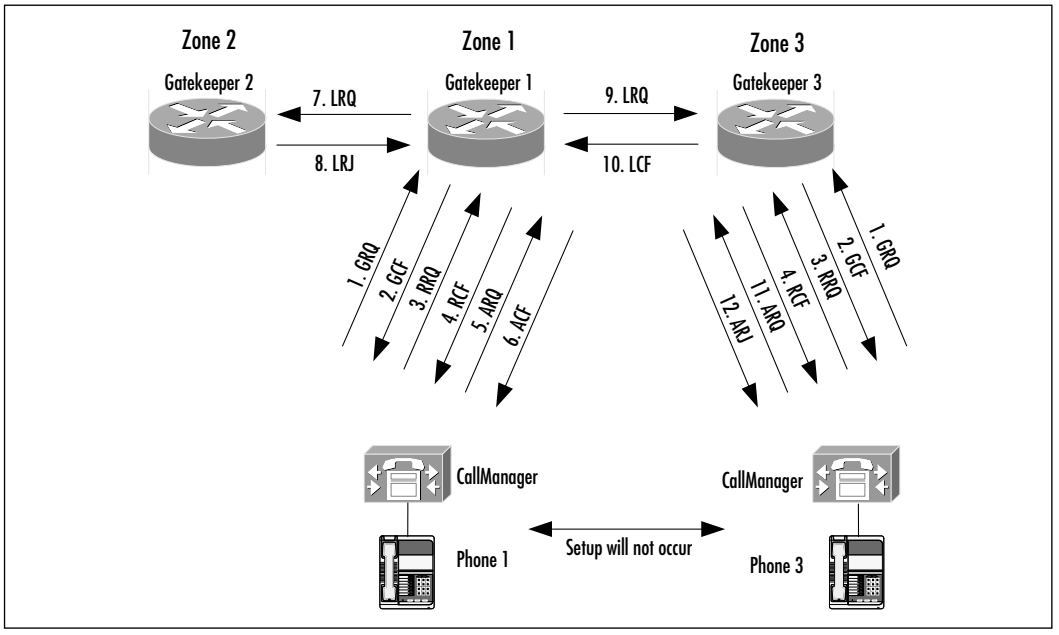


Figure 5.2 Gatekeeper Call Rejection



Both CallManagers again discover and register with their gatekeeper (Steps 1 to 4). When Phone 1 places a call, its CallManager sends an admission request to its gatekeeper to determine if it may place the call (Steps 5 to 6). Gatekeeper 1 uses a location request to determine if the endpoint is in zone 2. It receives a rejection from Gatekeeper 2 (Steps 7 to 8). This can happen for a variety of reasons. For example, it can occur if Gatekeeper 2 normally has a gateway that can service this type of call, but the gateway is currently down. It can also happen if Gatekeepers 2 and 3 both can service the call but Gatekeeper 2 has reached its limit on calls. Gatekeeper 1 then locates the endpoint in zone 3 (Steps 9 to 10). The CallManager receiving the call (on behalf of Phone 3) sends an admission request to its gatekeeper to determine if it may receive the call. Gatekeeper 3 rejects the call (Steps 11 to 12) because adding this call would exceed the bandwidth configured for H.323 calls for zone 3. This prevents the call from being set up over the IP network. If it is configured to do so, the CallManager will attempt to place this call over the PSTN.

Design Considerations

When designing your H.323 network, you should try to think of what a complete, fully implemented H.323 design would look like. Attempt to implement your (probably smaller) H.323 network with this vision in mind. Networks inevitably continuously grow. It is much easier to start with a large-scale design even in a small network and scale it to a large network than it is to begin with a small-scale design and scale it to a large network.

Although the gatekeeper can be located anywhere in the network, since it is a central point of control, the optimal location is near the center of the network or near the center of the H.323 network. The gatekeeper should be connected to the network via a 10 Mbps or—ideally—a 100 Mbps Ethernet switched link.

Designing & Planning…

Using E.164 Numbers or H.323 IDs

When you deploy your H.323 network, you must identify endpoints and gateways either by E.164 numbers (telephone numbers) or H.323 IDs (text strings). Cisco's implementation requires H.323 IDs use an e-mail address format (user@company.com).

Continued

While the latter format may seem appealing, since each of your users will have a unique e-mail address, in practice it is rarely used. Commonly E.164 addressing is used in H.323 networks. One reason is people are used to dialing E.164 numbers for voice and video calls, not e-mail addresses. Another reason is E.164 addressing typically leads to a more organized, hierarchical addressing system. Most companies already have a telephone addressing system in place. You can configure your H.323 similarly to your existing dialing plan. For example if your company uses four-digit telephone extensions, your existing and H.323 dialing plan might look like that shown in Table 5.2.

Table 5.2 Sample H.323 Dialing Plan

H.323 Configured Dial Pattern	Meaning of Dial Pattern	How the Call Is Routed
9*	Dial 9 and any number of digits	The call is directed to the PSTN (outside call).
5….	Dial 5 and any four digits	The five-digit number is routed to the Chicago office.
4….	Dial 4 and any four digits	The five-digit number is routed to the Dallas office.
3….	Dial 3 and any four digits	The five-digit number is routed to the Atlanta office.
8*	Dial 8 and any number of digits	The call is a video conference call and routed to a video gateway.

Using this plan, you would probably install gateways that use T1 lines to access the PSTN via a local carrier. Calls beginning with "9" would get routed to these gateways. You might use gateways that create Voice over IP (VoIP) sessions for calls to the Chicago, Dallas, or Atlanta offices (using the internal data network). Calls beginning with "3, 4, or 5" would get routed to these gateways. You might install videoconferencing gateways that could complete videoconferencing calls. Calls beginning with "8" would get routed to these gateways.

Using Bandwidth Limits in Your Network

As discussed earlier, one of the most useful features of a gatekeeper is bandwidth control—being able to efficiently utilize your WAN bandwidth while leaving

sufficient bandwidth for other applications. Since the gatekeeper can act as a "bandwidth policeman," it can save money by allowing calls to use internal WAN circuits rather than the traditional PSTN. Yet it is also intelligent enough to limit the number of calls based on the bandwidth limits you configure.

There are several ways to limit bandwidth utilization in your network. For example, the gatekeeper can limit the bandwidth used by any given session (voice or video call). If you are not yet prepared to support video (or any other high-bandwidth) calls in your network, you could limit the bandwidth of a session to the bandwidth used by a voice call. Thus, if you use g.711 CODECs in your network, calls use approximately 80 Kbps (64 Kbps of data plus some IP overhead). You could configure the following command to limit the bandwidth used by any call to 80 Kbps:

```
bandwidth session default 80
```

To limit the total amount of H.323 traffic—both within a zone and to and from that zone from other zones—use the *total* keyword. This command limits the total H.323 traffic for all zones to 1 Mbps:

```
bandwidth total default 1000
```

To limit the bandwidth to and from a particular zone, use the *interzone* keyword. Meanwhile, to limit bandwidth into and out of the sales zone to 512 Kbps, use:

```
bandwidth interzone zone sales 512
```

Using Accounting within Your Network

Accounting information about traffic flows is always useful. Where the traffic is flowing, when the traffic is flowing, and how much traffic is flowing are all useful information a gatekeeper can log with a TACACS+ or RADIUS server. This data can aid in capacity planning as well as in optimizing the network.

To enable Administration, Authorization, and Authentication (AAA) accounting and define a RADIUS server, use the following commands:

```
aaa new-model
radius-server host 192.168.51.51
radius-server key 0 (password)
```

To configure the gatekeeper to perform accounting, use the following commands:

```
aaa accounting connection h323 start-stop group radius
gatekeeper
  aaa accounting
```

The *start-stop* keyword issues an AAA record to the RADIUS server when a call starts and when it stops. The *wait-start* keyword can be used in place of this. With this configuration, the call does not start until the server has received the start record. This can introduce delays in the call being placed, but it guarantees that all accounting records are in place before any call is started. This should only be used when accurate accounting records are of the utmost importance.

Using Multicast or Unicast Addresses to Locate the Gatekeeper

When configuring endpoints and gateways to use gatekeepers, either a multicast or unicast address must be used to locate and register with the gatekeeper. When using multicast addressing, there is less control over where endpoints will register since they will register with the first gatekeeper to reply to their request. If the network is configured such that only one gatekeeper will register any given endpoint or gateway, there is very little advantage to using multicast addressing.

Multicast addressing does allow the address of the gatekeeper to change with no reconfiguration on endpoints or gateways, though with unicast addressing, using a DNS name rather than a specific IP address accomplishes the same goal. Using multicast addressing does require multicast routing to be enabled in the network.

Designing a Large H.323 Network

To achieve stability in a large H.323 network, a two-tiered gatekeeper design is often employed. This is sometimes called a "directory of gatekeepers" approach since one gatekeeper acts strictly as a gatekeeper for all other gatekeepers.

In this design, two levels of gatekeeper are introduced. The *lower* level gatekeepers are traditional gatekeepers: they manage zones, calls, and register endpoints and gateways. However, rather than being configured with all other gatekeepers in the network (and all their zones), they are configured to use a single gatekeeper. This gatekeeper is the *upper* level gatekeeper. It is a special gatekeeper in that it does not register any endpoints or manage any zones. Instead, it is configured with all other gatekeepers and all of the zones (i.e., E.164 prefixes) that they serve. In this sense this gatekeeper acts as a "directory" of zones. Its primary function is to assist gatekeepers in locating the correct gatekeeper for any given endpoint.

This design removes the many-to-many gatekeeper configurations required in a flat or single level design, and creates a hierarchical design where all lower level gatekeepers lead to a single point of control. This design also makes growing an H.323 network much easier. When new zones and a new gatekeeper are added, the only change to the existing network is to add the correct information to the "directory" gatekeeper. In the flat design, every gatekeeper must be updated with information about the new gatekeeper.

Note that this hierarchical structure is a logical design of gatekeepers. The underlying network should provide connectivity between endpoints with the fewest number of hops possible (while still providing a structured network design). Adding several additional hops between endpoints can contribute to slower and lower quality voice and video performance.

NOTE

As of 12.1(5)XM, the upper level, or directory gatekeeper could only service approximately six lower level gatekeepers. As this limit will likely change often, you should check with your local Cisco resource or the Cisco TAC for updated limits.

Zone Designs

To create an optimum H.323 zone design, it is important to understand what zones are, what functions they perform and how calls are routed between zones. Think of designing an IP network: it is critical to understand what subnets are and how packets are routed between subnets. Although H.323 zones are typically much larger than one IP subnet, the same type of understanding is required.

Zones are simply collections of endpoints, gateways and Multipoint Control Units (MCUs—provide H.323 conferencing of three or more devices). They can be grouped in any manner that makes administration and organization easier. You can use one large zone or many small zones, though an eye should be kept on future scalability. One gatekeeper services each zone, though a gatekeeper can service multiple zones. You create zones by:

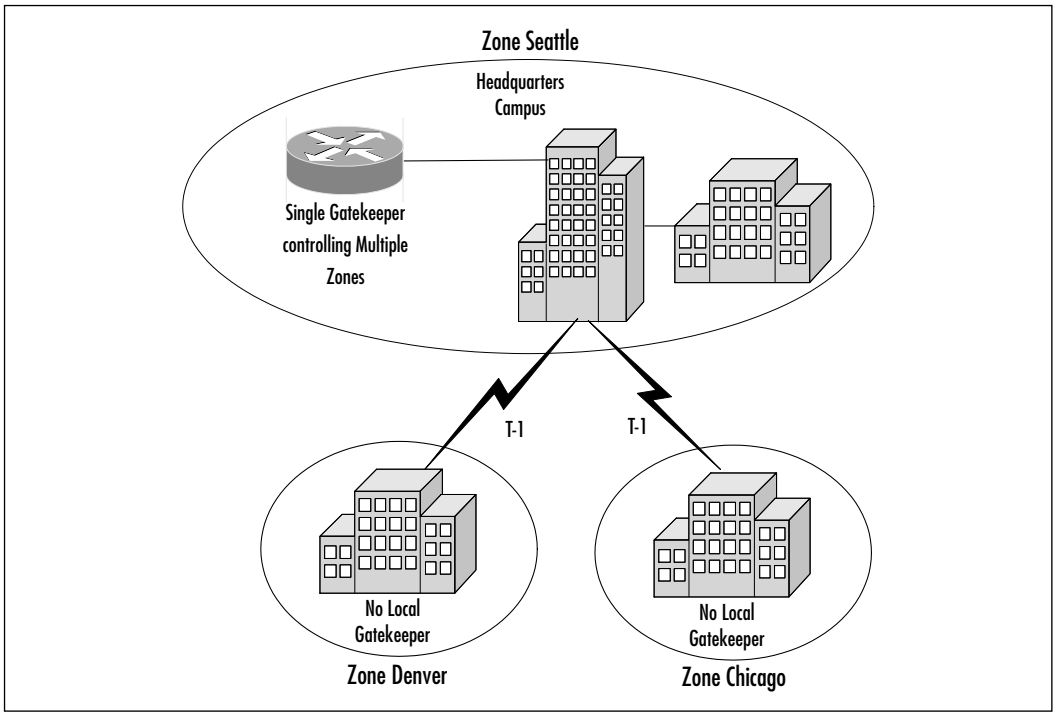
- Configuring each gatekeeper to place subnets into specific zones
- Configuring endpoints and gateways to register with specific gatekeepers (based on the subnets they are in)

To accomplish the former task, you should use the **zone subnet** command to define what subnets the gatekeeper will service. A gatekeeper can service many subnets, and any requests from subnets not configured on a gatekeeper will be rejected. Using this command allows you to control the gatekeeper that endpoints and CallManagers use for registration. This should be done in all cases, though this is critical if you are using multicast since with multicast an endpoint will find all available gatekeepers and register with the first one to respond. If more than one gatekeeper is configured to service the same subnet, the endpoint may register with a gatekeeper to which you did not intend it to register.

Implementing Zones in Your Network

When deploying H.323 zones in your network, consider making each site connected by a WAN link its own zone. One primary advantage of this design is that each zone can be configured to use a specified amount of bandwidth for voice and video calls. This is useful since the WAN is almost always where bandwidth will be limited, and bandwidth limits can be set on a per-zone basis. This approach is shown in Figure 5.3.

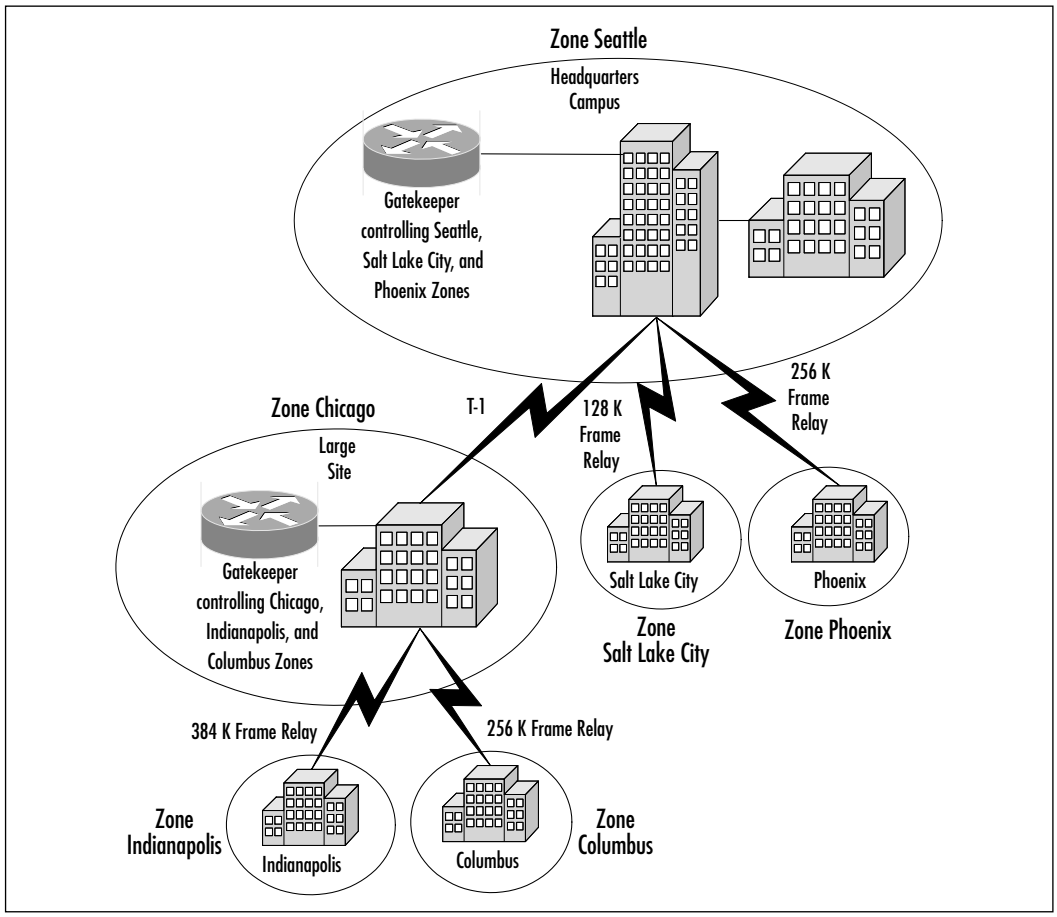
Figure 5.3 Multiple Zones with a Single Gatekeeper



For example, assume the Denver zone has a significant amount of time-sensitive traffic other than voice and video calls. For this zone, you would probably want to limit the voice and video calls to only 512 Kbps or 768 Kbps to leave sufficient bandwidth for the other time-sensitive applications. However, assume the Chicago zone does not have a significant amount of time-sensitive traffic. For this zone, you would probably want to allow the voice and video calls to consume much more bandwidth, such as 1024 Kbps or 1280 Kbps. Allowing more calls would save costs (versus using the PSTN) but would still provide some bandwidth for all other applications. This configuration for all three zones can be placed in the Seattle gatekeeper.

As you can see in Figure 5.4, remote sites can have their own gatekeeper, though they are not required to.

Figure 5.4 Multiple Zones with Multiple Gatekeepers



Small offices, such as sales offices do not need their own gatekeepers. Especially if there are a large number of small offices, assigning each their own gatekeeper becomes unmanageable and expensive. Instead small offices can be placed into their own zones, but can utilize a remote gatekeeper. The gatekeeper at the major site can manage the zone at that site and zones for all smaller sites that connect to that site. The gatekeeper limits the amount of bandwidth used by calls into or out of each zone.

As shown in Figure 5.4, the Seattle gatekeeper can manage the Seattle, Salt Lake City, and Phoenix zones. Placing each site in its own zone allows each site (zone) to set its own limit on voice and video traffic. This is important since different sites will have different size WAN connections as well as different traffic patterns.

Chicago is a large enough site to warrant its own gatekeeper. The Chicago gatekeeper can manage the Chicago, Indianapolis, and Columbus zones. It will control how many voice and video calls are placed to and from Chicago, to and from Indianapolis, and to and from Columbus.

If a call is placed from Phoenix to Columbus, the Seattle gatekeeper determines whether the Phoenix endpoint can place the call. The Chicago gatekeeper, meanwhile, determines whether the Columbus endpoint can receive it.

Alternate Zone Designs

An alternative method of creating zones is to create them functionally, rather than geographically. You can configure zones with different security restrictions and use the **zone subnet** command to only allow users to join their appropriate zone.

Each user (endpoint) can be authenticated before being admitted to their zone. This can be done with endpoint authentication. Version 1 of H.323 does not have comprehensive authentication. Users must piggyback their password onto their H.323 registration with a predefined password separator character separating the two. The gatekeeper can then collect the password and authenticate it to the RADIUS or TACACS+ server. While this approach is less common and far more complex than the geographical approach, it does increase security. It is typically used when security, not bandwidth utilization, is the primary concern of the H.323 network.

Routing Calls between Zones

Call routing is based on either e-mail addresses (H.323 IDs—a string) or E.164 telephone numbers. However, E.164 gives you more flexibility. You can assign

zones by area code, country code, area code + local exchange, or basically anything you want.

When gatekeepers are attempting to route calls (resolve the endpoint address of a call) they determine the destination zone using their zone prefix configuration. This will allow the gatekeeper to locate the correct zone and the associated gatekeeper. If the endpoint is part of a remote zone the gatekeeper will send a location request to the appropriate gatekeeper to determine how to resolve the address. If the endpoint is part of a local zone (or if the gatekeeper is servicing an incoming call to one of its local zones), the gatekeeper will either find the exact endpoint (which has registered with them) or an appropriate gateway.

If the gatekeeper needs to route a call to a gateway, it will select a gateway based on the technology prefix the gateway used when it registered with the gatekeeper. For example, when gateways register with the gatekeeper they typically will register with one or more technology prefixes that they support. Perhaps the technology prefix 1# is used to designate voice gateways while 2# is used for ISDN gateways. When the gatekeeper receives a call request beginning with 1# to one of its local zones, it will select one gateway from the pool of gateways that have registered with that technology prefix. This requires callers to dial the appropriate technology prefix based on the type of call (voice, ISDN, and so on) they are placing.

On the gatekeeper, you can use the **gw-type-prefix** command for several purposes. First, if a gateway is incapable of registering a technology prefix (such as 1#) you can use this command with the *gw ipaddr* keywords to manually define the technology prefix a gateway supports.

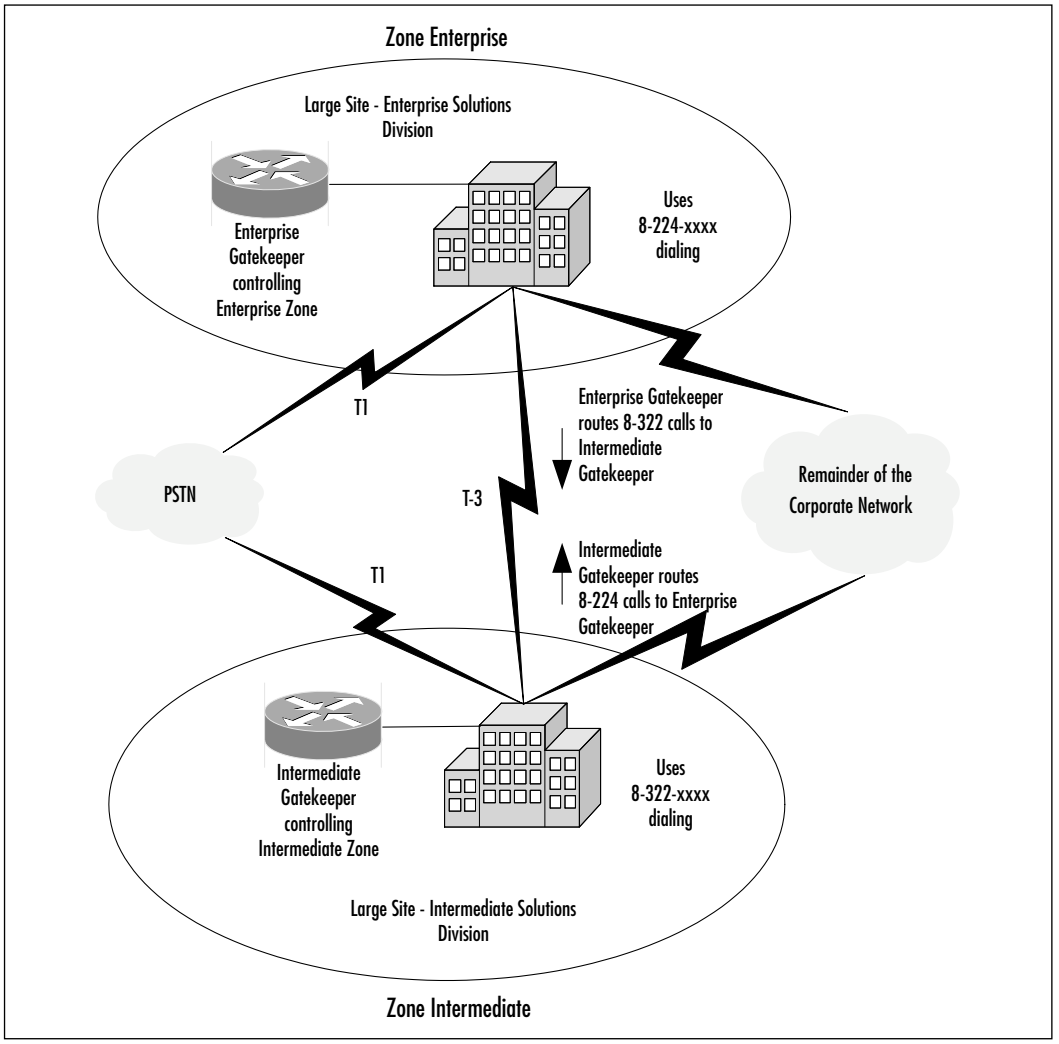
Second, you can use the **gw-type-prefix** command to define a default technology prefix to be used if a call does not have a technology prefix. Since voice calls are so common, many organizations define voice gateways to be the default technology and thus do not require callers to use any technology prefix when placing voice calls. In the previous example, you could define 1# (voice calls) to be the default technology prefix. Thus callers would not need to dial a preceding 1# for voice calls. Those calls would automatically get sent to the default technology gateways: the voice gateways. Callers would only need to use a technology prefix (2#) for ISDN calls.

Finally, the **gw-type-prefix** command can be used with the *hopoff* keyword to force calls with a certain type of technology prefix to route through a particular gatekeeper regardless of any **zone prefix** commands. For example, you may have many zone prefixes defined to enable routing of voice calls. However, you may want to route ISDN calls (a different technology prefix than voice calls) to a

particular gatekeeper (i.e., where the ISDN gateways reside) regardless of whether the dialed number matches a **zone prefix** command.

Figure 5.5 shows two locations of a large international corporation. Only two of the many locations within the corporation are shown. One site specializes in large, enterprise products while the other produces smaller, intermediate products.

Figure 5.5 Interzone Routing



The company has implemented a dial plan where you dial a "9" to get an outside line. For calls within the company, you use an "8" followed by seven digits. The Enterprise group uses the 224-xxxx exchange, while the Intermediate

group uses the 322-xxxx exchange. These patterns will be used within the gatekeepers for routing calls between zones. Here is how you would configure each gatekeeper.

For the Enterprise Gatekeeper:

```
zone local enterprise company.com
zone remote intermediate company.com 172.16.128.1
zone prefix enterprise 8224....
zone prefix intermediate 8322....
```

For the Intermediate Gatekeeper:

```
zone local intermediate company.com
zone remote enterprise company.com 172.16.192.1
zone prefix intermediate 8322....
zone prefix enterprise 8224....
```

Configuring & Implementing…

Implementing Multiple Gatekeepers

When determining how many gatekeepers to implement in a network, you should examine whether your current network is mostly centralized or more distributed. If your company has one large campus for its headquarters and only small remote offices, this is a mostly *centralized* network. If your company has a headquarters location, but several large remote facilities, this is a *distributed* network.

A centralized network can probably be implemented with a single gatekeeper, given the gatekeeper is placed in a central location. A distributed network is probably best served by several gatekeepers. Often each large location will be its own zone, maintained with its own gatekeeper. Whether a site is large enough to warrant its own gatekeeper is partially a function of how many users are there and how many H.323 endpoints and gateways there are. The number of H.323 endpoints and gateways is important since this directly affects the amount of activity on the H.323 network. The number of users at a site is indirectly important since this can affect how large the H.323 network may eventually grow.

A simple yet effective design is to place one gatekeeper at each of your company's major sites (again, this is a subjective decision). Each of

Continued

the gatekeepers will maintain that zone and any zones that connect via a WAN connection to that site. Gateways and endpoints (CallManagers and so on) will join the correct zone via the configuration on the endpoint and the configuration on the gatekeeper.

For example, CallManager allows you to configure the IP address of the gatekeeper in the *Gatekeeper Name* field. This controls the gatekeeper with which the CallManager will register. The gatekeeper is configured to place certain IP ranges into specific zones. For example, the commands that follow place 10.10.10.10 (a CallManager, perhaps) into the zone *engineering*.

```
zone local engineering company.com
zone subnet engineering 10.10.10.10/32 enable
```

Alternatively you could configure the gatekeeper to allow the entire subnet to join the zone engineering.

```
zone local engineering company.com
zone subnet engineering 10.10.10.0/24 enable
```

Each gatekeeper will be aware of all other gatekeepers. A structured dial plan will be established so that each gatekeeper knows how to route calls based on the E.164 or technology prefix.

If your company is just starting to deploy an H.323 network, you can probably deploy the gatekeeper(s) on an existing router. In this case, add a new subnet to the gatekeeper and configure the router's new IP address as a secondary IP address or as a new loopback. Use this as the address used by endpoints and CallManagers to register with the gatekeeper. If the H.323 network grows large enough to warrant a separate router as the gatekeeper, simply move this IP address to the new gatekeeper router. That way all your endpoints can still register with the same IP address.

A Gatekeeper's Role in Voice and Video Networking

A gatekeeper plays a key role in a voice network. It translates addresses so that when a user dials a telephone number, the gatekeeper determines the IP address associated with it. The gatekeeper will admit endpoints and calls into the network based on configured parameters. The gatekeeper can also provide authentication and accounting of all calls placed in the network.

Even with video calls, users are accustomed to dialing phone numbers to complete calls. It is the gatekeeper's responsibility to resolve the dialed phone number to the correct IP address. In video networks, admission control and call control are critical functions.

Admission control is important because sensitive data is often kept in video presentations, such as new product plans or financial announcements. It is vital that a gatekeeper monitor the endpoints that have access to this content.

Call control is important because video usually requires more bandwidth than voice calls. For this reason, the gatekeeper must closely monitor call admission to assure both that the network has adequate bandwidth to provide a quality call and that the call does not consume so much bandwidth that other applications are ineffective.

Choosing a Gatekeeper Platform

The exact gatekeeper platform required for an H.323 network depends on how large the H.323 network is and how many other functions (if any) the router will be performing. The more endpoints, gateways, and calls in an H.323 network, the larger the router platform required for the gatekeeper. A router dedicated for gatekeeper tasks will have considerably more resources available than a router performing several other functions in addition to gatekeeper.

Ideally, the gatekeeper should connect to the network using 100 Mbps Ethernet, though for small networks, a 10 Mbps Ethernet connection will suffice.

When selecting router memory, remember the rule: There is no such thing as too much memory. However, it should be noted that gatekeeper functionality does not require an enormous amount of memory. For example, to support the configuration for 10,000 zones, only an additional 4MB of memory is required. More memory would be necessary to monitor the calls, but this gives you an idea that the gatekeeper does not require tremendous memory.

Selecting a Router Hardware Platform

Although any of the supported platforms will run gatekeeper, Cisco recommends the 2600, 3600, and 7200 platforms. The 3600 and 7200 will provide the most powerful gatekeeper implementations. The information in Table 5.3, Gatekeeper Platform Performance Statistics, is provided by Cisco to allow users to estimate the router gatekeeper required for their network.

Table 5.3 Gatekeeper Hardware Platform Statistics

Gatekeeper Platform	Memory	Maximum Calls per Second for Approximately 50 percent CPU Utilization
Cisco 2600	56MB	7
Cisco 3620	56MB	10
Cisco 3640	128MB	24
Cisco 3660	256MB	35
Cisco 7200/NPE300	256MB	50

Although it will likely be difficult to estimate the number of calls per second that will occur in your network, this table can be used in a more relative way. That is, going from a 2600 to a 3620 increases gatekeeper performance by approximately 50 percent. Likewise, going to a 3640 with 128MB of memory increases gatekeeper performance by more than 100 percent compared to a 3620 with 56MB of memory, and so on.

If the H.323 network is being deployed as an organized project where funding is available for the initial purchase but not necessarily for follow-on purchases, it may be safer to jump to a 3600 router as the gatekeeper. If the H.323 network is being deployed where funding is limited, a spare or lightly loaded 2600 should suffice as your gatekeeper.

Selecting an IOS

Although the gatekeeper functionality was introduced in some versions of 12.0, significant enhancements and additional functionality have been made in 12.1 and 12.2. At a minimum, a recent fix version of 12.1 should be used. If possible, the latest fix version of 12.2 or later IOS should be used.

Redundancy

Every network has different redundancy requirements, from total redundancy for every element to no redundancy required. There are several different ways redundancy can be incorporated into your gatekeeper design. The following sections outline the most common methods. As discussed next, the most common (and one of the most effective) ways to achieve redundancy is via Hot Standby Router Protocol (HSRP).

Configuring HSRP between Gatekeepers

Cisco gatekeepers run a standard version IOS with the H.323/MCM feature set. Thus, they have many common IOS capabilities, including HSRP. As with any routing implementation, HSRP is an excellent way to provide redundancy between routers.

In this scenario, you will probably want to configure the endpoints (such as CallManagers) and other gatekeepers to register with a specific IP address (which will be required anyway unless you plan to implement the multicast solution). In this case, configure your endpoints and other gatekeepers to register with the HSRP address. This way, regardless of which router is the active gatekeeper, all devices will be able to successfully register with the gatekeeper.

You should configure the gatekeepers to use the HSRP address as their local RAS address for all zones. You can do this by using the **zone local** command and specifying at the end of the command a local IP address that the gatekeeper should use. This will force the gatekeepers to use that address for all communication with endpoints and gateways in that zone.

In the event of a failure where the back-up gatekeeper becomes active, failover will not be transparent. The two HSRP gatekeepers do not share state tables, thus all endpoints and other gatekeepers will need to re-register with the newly active gatekeeper. This should only cause a minor service disruption.

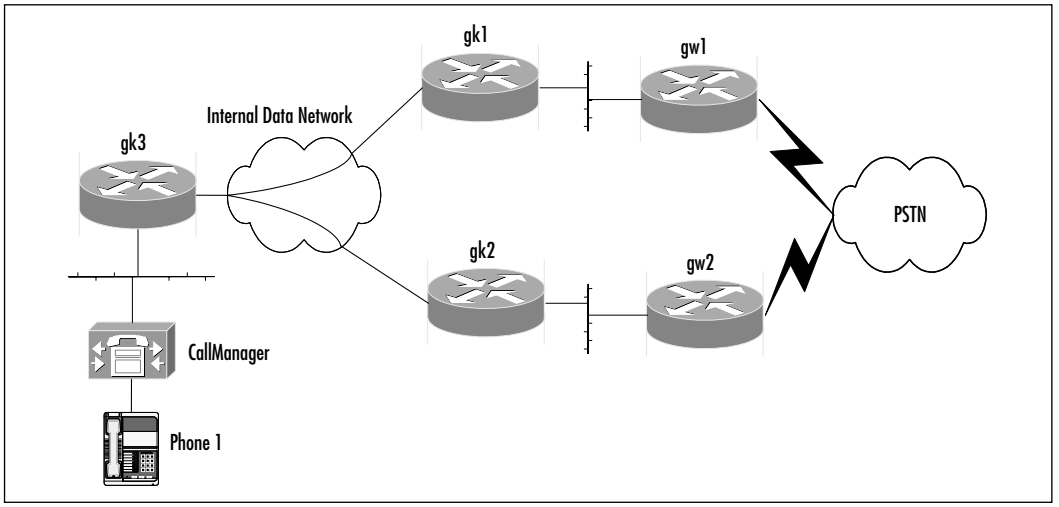
In order to assure full functionality in the event of a failover, maintain exact gatekeeper configurations on both devices. When changes are made, be diligent to make identical changes on the backup router.

You can implement HSRP with two dedicated routers, though often this is not economically practical. A more feasible approach might be to deploy the primary gatekeeper with either a dedicated router or a router that does not have a heavy load (such as a router that connects to a test environment, for example). The backup gatekeeper can be an existing router that performs several other functions. Although overall performance may be degraded in a failure scenario, this will likely occur very infrequently. In this case, you may want to configure the primary gatekeeper with not only a higher HSRP *priority* but also with HSRP *preempt* so that it may resume the gatekeeper function as soon as it has recovered. Use Table 5.3 to gauge the approximate size of the router required for both your primary and HSRP standby router.

Using Technology Prefixes for Redundancy

Using technology prefixes, you can configure more than one remote gatekeeper to which the local gatekeeper can use for placing calls. As shown in Figure 5.6, you can install two gateways that both place calls to the PSTN (gw1 and gw2). Calls to the PSTN can be defined as technology prefix 1#. For redundancy, these gateways can register using technology prefix 1# with two different gatekeepers, gk1 and gk2. Now both gatekeepers, gk1 and gk2, can route calls to the PSTN. Gk1 and gk2 can be in the same zone or different zones. Gatekeeper 3, gk3, can be defined to use gk1 and gk2 to place calls to the PSTN (via technology prefix 1#). This provides redundancy for both the gatekeepers and the gateways.

Figure 5.6 Adding Redundancy to Gatekeeper Designs



For technology prefix 1#, gk3 can be configured to prefer one gatekeeper (*sequential*) or to request call placement to both, using whichever gatekeeper responds first (*blast*). In sequential mode, gatekeepers will be tried in the order they are listed in the **gw-type-prefix** command.

For further redundancy, you can configure the two gateways to register with either gatekeeper. The gateways attempt to register with gatekeepers in the order in which they are configured. If their registration fails, they will try the second gatekeeper, and so on. You could configure gw1 to register with gk1, then gk2. You could configure gw2 to register with gk2, then gk1. This would allow both gateways to be operational even if one of the gatekeepers failed. In this approach, it may be simpler to place gk1, gk2, gw1, and gw2 in the same zone.

Using Zone Prefixes and Gatekeeper Clusters for Redundancy

The **zone prefix** command does not allow you to list multiple destination gatekeepers. Gatekeeper clusters avoid this limitation and allow gatekeepers to act in a redundant manner.

In Figure 5.6, gk1 and gk2 can be grouped into a cluster. For example, gk1 and gk2 could both be defined in zone "zoneone" and cluster "cluster1." Gk1 and gk2 would list the opposite gatekeeper as belonging to the cluster. They can be configured to transfer calls to the alternate gatekeeper based on many parameters, such as CPU or memory utilization. In this example, gk1 uses a limit of 100 active calls at which point calls are transferred to the alternate:

```
gatekeeper
  zone local zoneone company.com
  zone cluster local cluster1 zoneone
    element gk2's IP address
    exit
  load balance calls 100
```

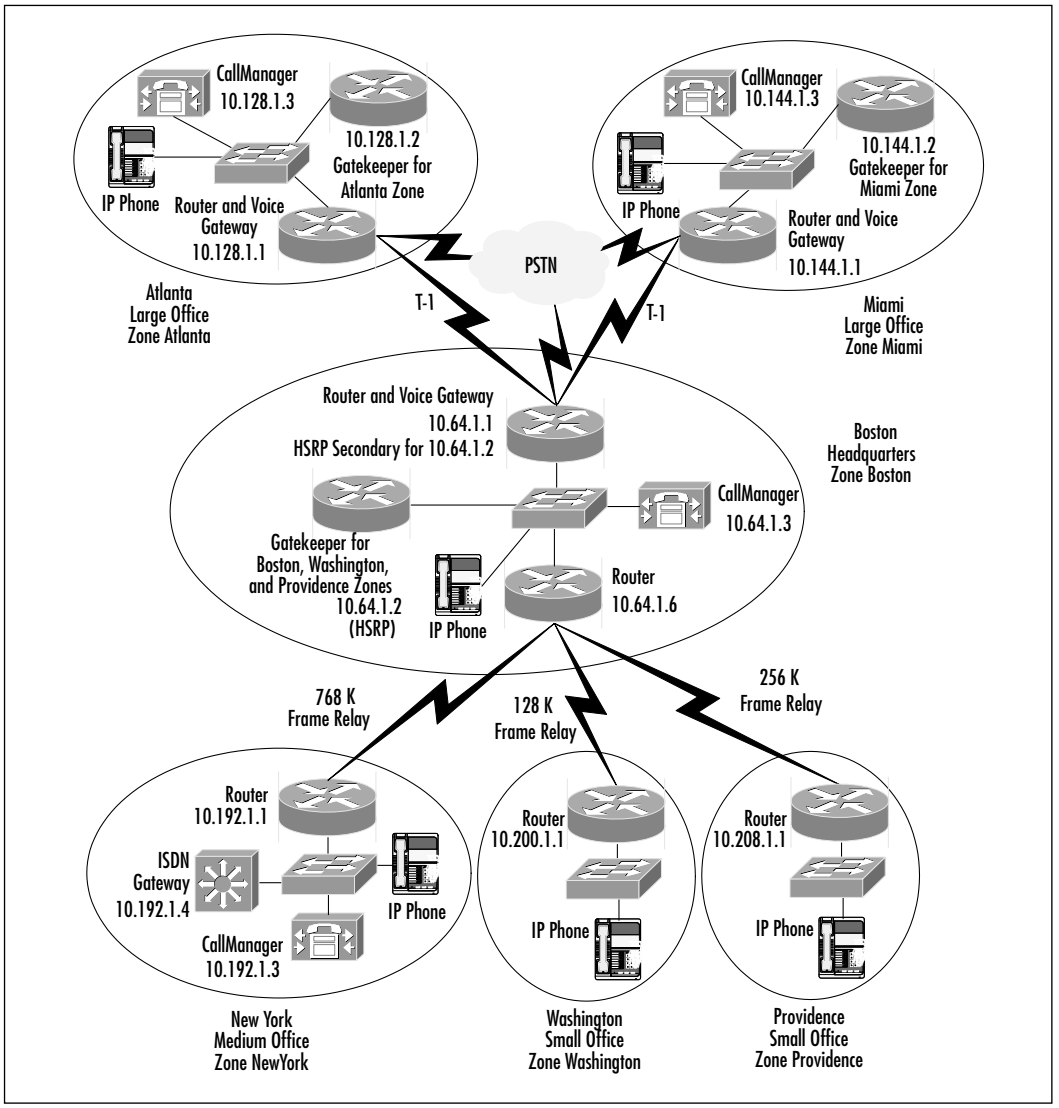
The remote cluster (cluster1) and each of the gatekeepers in the cluster are defined on gk3. Then the appropriate zone prefix is configured to use the cluster. Continuing with this example, gk3 is configured with cluster "cluster1," of which gk1 and gk2 are both members. Gk3 is then configured to route calls for the 312 area code to cluster1:

```
gatekeeper
  zone local zonetwo company.com
  zone cluster remote cluster1 company.com
    element gk1 gk1's IP address
    element gk2 gk2's IP address
    exit
  zone prefix cluster1 312.......
```

Placing and Configuring Gatekeepers: A Case Study

This case study will examine the network design and configuration of a large manufacturing company that has deployed a voice and video network; it will focus primarily on the gatekeeper placement and configuration. The company has six sites, as shown in Figure 5.7.

Figure 5.7 Case Study: Company Network



The company has decided to place gatekeepers in their large offices (Boston, Atlanta, and Miami), but not their small or medium offices (New York, Washington, and Providence). Deploying gatekeepers in the larger offices will offload call admission and bandwidth management from the headquarters (Boston) gatekeeper to a local gatekeeper. Likewise, the medium and large sites will have their own CallManager while the two small offices will use the Boston CallManager. Since the New York office makes a large amount of videoconferences, that site has an ISDN gateway to make videoconferences off of the corporate network. This gateway is available for any site that needs to make a videoconference with another company.

Configuring Local Zones

The first step in configuring the gatekeepers is to define the zones the gatekeeper will manage. These zones are known as *local* zones. Even though they may be located at another site (and thus not "local" in a traditional sense) all zones managed by a gatekeeper are considered local to that gatekeeper. The following shows the commands for each gatekeeper in gatekeeper configuration mode.

For Atlanta:

```
zone local atlanta company.com
```

For Boston:

```
zone local boston company.com 10.64.1.2
zone local newyork company.com
zone local washington company.com
zone local providence company.com
```

For Miami:

```
zone local miami company.com
```

Configuring the Zone Subnet

The next step is to configure the gatekeepers to accept each CallManager that will register with them. This configuration will not only allow each CallManager to register, but it will also configure the zone to which the CallManager is a member. The following shows the commands for each gatekeeper in gatekeeper configuration mode.

For Atlanta:

```
no zone subnet atlanta 0.0.0.0/0 enable
zone local subnet atlanta 10.128.1.3/32 enable
```

For Boston:

```
no zone subnet boston 0.0.0.0/0 enable
zone local subnet boston 10.64.1.3/32 enable
no zone subnet newyork 0.0.0.0/0 enable
zone local subnet newyork 10.192.1.3/32 enable
```

For Miami:

```
no zone subnet miami 0.0.0.0/0 enable
zone local subnet miami 10.144.1.3/32 enable
```

NOTE

The **no zone subnet 0.0.0.0/0** command is required since, by default, the gatekeeper will accept any registrations. This command stops that behavior and allows each CallManager to be accepted and placed in its zone.

Configuring Zone Bandwidth

The next step is to define the amount of bandwidth you will allow to be used by voice and video calls for each WAN link (or zone). The company does not have any unusual bandwidth requirements, so we will allow 75 percent of each link's bandwidth to be used by voice and video calls. The bandwidth specified by the **bandwidth** command is in kilobits/second.

For Atlanta:

```
bandwidth interzone zone atlanta 1152
```

For Boston:

```
bandwidth interzone zone newyork 384
bandwidth interzone zone washington 96
bandwidth interzone zone providence 192
```

For Miami:

```
bandwidth interzone zone miami 1152
```

Configuring Remote Zones

The next step is to configure all of the remote zones on each gatekeeper so each gatekeeper is aware of every zone in the network. Again, *remote* zone simply refers to any H.323 zone managed by another gatekeeper. The IP address designates the gatekeeper for that zone, so the local gatekeeper will know how to contact the gatekeeper for each zone.

For Atlanta:

```
zone remote miami company.com 10.144.1.2
zone remote boston company.com 10.64.1.2
zone remote newyork company.com 10.64.1.2
zone remote washington company.com 10.64.1.2
zone remote providence company.com 10.64.1.2
```

For Boston:

```
zone remote atlanta company.com 10.128.1.2
zone remote miami company.com 10.144.1.2
```

For Miami:

```
zone remote atlanta company.com 10.128.1.2
zone remote boston company.com 10.64.1.2
zone remote newyork company.com 10.64.1.2
zone remote washington company.com 10.64.1.2
zone remote providence company.com 10.64.1.2
```

Configuring the Dial Plan

The company uses a dial plan as shown in Table 5.4.

Table 5.4 Case Study: Company Dial Plan

H.323 Configured Dial Pattern	Meaning of Dial Pattern	How the Call Is Routed
9*	Dial 9 and any number of digits	The call is directed to the PSTN via the nearest location (Boston, Atlanta, or Miami).
8#*	Dial 8, then #, then any number of digits	The call is directed to the ISDN gateway in New York.
1…	Dial 1 and any three digits	The four-digit number is routed to the Boston office.
2…	Dial 2 and any three digits	The four-digit number is routed to the Atlanta office.
3…	Dial 3 and any three digits	The four-digit number is routed to the Miami office.
4…	Dial 4 and any three digits	The four-digit number is routed to the New York office.
5…	Dial 5 and any three digits	The four-digit number is routed to the Washington office.
6...	Dial 6 and any three digits	The four-digit number is routed to the Providence office.

Each gatekeeper must be configured according to the dial plan to be able to route calls to the correct zone. Each gatekeeper is configured identically with each of the dial prefixes shown in the preceding table so it is able to route any call.

For Atlanta:

```
zone prefix boston 1...
zone prefix atlanta 2...
zone prefix miami 3...
zone prefix newyork 4...
zone prefix washington 5...
zone prefix providence 6...
```

For Boston:

```
zone prefix boston 1...
zone prefix atlanta 2...
zone prefix miami 3...
```

```
zone prefix newyork 4...
zone prefix washington 5...
zone prefix providence 6...
```

For Miami:

```
zone prefix boston 1...
zone prefix atlanta 2...
zone prefix miami 3...
zone prefix newyork 4...
zone prefix washington 5...
zone prefix providence 6...
```

Configuring Gateway Type

Next, the gatekeepers must be configured to route calls destined for outside the company network (ISDN videoconference calls and local and long distance voice calls). The technology prefix 8# will be used for ISDN calls via the ISDN gateway in New York. Thus, Atlanta and Miami will route these calls to the Boston gatekeeper (which manages the newyork zone).

The technology prefix 7# will be used for calls for which the gatekeeper has no other routing defined. Since the prefixes defined in the previous section cover voice calls within the company and the 8# technology prefix cover ISDN calls, the 7# technology prefix will cover all voice calls destined for outside of the company network.

For Atlanta:

```
gw-type-prefix 8# hopoff boston
gw-type-prefix 7#* default-technology
```

For Boston:

```
gw-type prefix 8# gw ipaddr 10.192.1.4
gw-type-prefix 7#* default-technology
```

For Miami:

```
gw-type-prefix 8# hopoff boston
gw-type-prefix 7#* default-technology
```

Note that the Boston gatekeeper must manually define the ISDN gateway's technology prefix and IP address because that gateway is not capable of registering its technology prefix automatically. The Atlanta and Miami gatekeepers define that 8# calls be sent to the Boston gatekeeper to force ISDN calls to the ISDN gateway in the newyork zone (controlled by the Boston gatekeeper).

NOTE

Each CallManager should register with its gatekeeper using the GateKeeperSupportedPrefix configuration. This will automatically register technology prefixes with the gatekeeper, and alleviates the need to manually define in the gatekeeper the technology prefixes supported by each CallManager. In the example in this section, each CallManager will register with the 7# technology prefix.

Configuring Gatekeeper HSRP

Additionally, the company wants to have some redundancy for the gatekeeper in Boston. The gatekeeper uses HSRP and acts as the primary for the 10.64.1.2 address. The WAN router (10.64.1.1) uses HSRP and acts as the secondary for the 10.64.1.2 address. This router needs the full gatekeeper configuration, but does not use it unless the primary gatekeeper fails. The HSRP configuration for these routers is as follows.

For the Boston Gatekeeper (10.64.1.2):

```
interface fast ethernet 0/0
  standby 1 ip 10.64.1.2
  standby 1 priority 105
  standby 1 preempt
```

For the Boston WAN Router (10.64.1.1):

```
interface fast ethernet 0/0
  standby 1 ip 10.64.1.2
```

Following the Call Flow

The Atlanta CallManager is configured to use the Atlanta Gatekeeper for call admission control. Thus when a user in Atlanta dials 1466, the Atlanta CallManager places a call request to the Atlanta gatekeeper. The CallManager knows that g.711 CODECs are used, so it requests an 80 Kbps call to the Atlanta gatekeeper. The Atlanta gatekeeper checks active calls and determines that if this call is placed, the total bandwidth used will be less than the defined interzone bandwidth of 1152 Kbps, thus the call is permitted. The Atlanta gatekeeper sees in its zone prefix configuration that all four-digit calls beginning with 1 (1...) get routed to the Boston zone. The Atlanta gatekeeper uses its zone remote configuration to determine that the Boston zone is controlled by gatekeeper 10.64.1.2. The Atlanta gatekeeper places a location request to 10.64.1.2 for destination 1466.

The Boston gatekeeper receives the request for location 1466. It uses its zone prefix configuration to also determine the destination zone is Boston. Since this is configured as a local zone, it attempts to find the endpoint or gateway for 1466, but because no endpoints have registered with this number, it uses its defined default technology prefix to route the call to any gateway that has registered a 7# technology prefix. The Boston CallManager has registered with a 7# technology prefix, so its IP address (10.64.1.3) is returned to the Atlanta gatekeeper, which returns it to the Atlanta CallManager.

The Atlanta CallManager attempts to place the call to the Boston CallManager. The Boston CallManager does not use a gatekeeper for call admission control. This is because CallManager can only register with the gatekeeper for one zone. Yet the Boston CallManager handles calls for three zones (Boston, Washington, and Providence). Since the CallManager registers with only one zone it would have no way to identify to a gatekeeper whether calls were actually being made from Boston, Washington, or Providence. Thus the Boston CallManager uses Location call admission control, which is configured directly on the CallManager. This method allows the CallManager to track each IP phone by its geographical location and monitor the number of calls (and thus bandwidth used) by each location. The Boston CallManager has sufficient bandwidth and the call is completed.

Washington and Providence are placed in separate zones so that the Boston gatekeeper can distinguish between them for other H.323 calls, such as videoconferences.

Summary

A gatekeeper provides much of the intelligence required to operate an efficient H.323 network. While not technically a requirement of an H.323 network, in practice the gatekeeper is extremely useful. It provides a single point of control for a zone—all the devices in its domain. Beyond providing address translation, admission control, and bandwidth control, the gatekeeper can provide functions such as accounting, call authorization, and directory services. In addition, gatekeepers manage devices making voice, video, and collaborative data calls.

The most common Cisco gatekeeper is the Multimedia Conference Manager (MCM). The MCM runs on a 2600, 3600, or 7200 router platform with the H.323/MCM IOS feature set.

Gatekeepers typically use E.164 addresses (telephone numbers) to route calls to the appropriate endpoint, gateway, or gatekeeper (for calls to remote zones). Small networks can be implemented with a single zone while larger, complex networks can use many zones.

Using the Hot Standby Router Protocol (HSRP) is an excellent way to provide gatekeeper redundancy in a network. Gateways can be configured to use either one of two different gatekeepers to provide redundancy.

Gatekeepers have the ability to not only limit endpoint participation in an H.323 network, they can control calls based on many factors, such as time of day, the number being called, or available bandwidth. These qualities make the gatekeeper an integral part of an H.323 network.

Solutions Fast Track

Understanding Gatekeeper Basics

- ☑ A gatekeeper is a central point of control for an H.323 (voice and video) network.
- ☑ Gatekeepers usually use E.164 addressing (telephone numbers) for identifying endpoints and routing calls within a network.
- ☑ Gatekeepers run an H.323/MCM feature set IOS on many common Cisco routers.

A Gatekeeper's Role in Voice and Video Networking

- ☑ Gatekeepers manage one or multiple zones and permit or reject calls into or out of each zone.
- ☑ Gatekeepers can provide accounting information for calls, such as length of call, time of call, number called, and so on.
- ☑ Cisco's Multimedia Conference Manager (MCM) can act as a proxy for increased security and QoS as well as a gatekeeper.
- ☑ Video gatekeepers can be embedded in the video controller or can be an MCM.
- ☑ Video gatekeepers interface with gateways for off-network calls, such as ISDN videoconferences.
- ☑ Gatekeepers monitor (and limit) bandwidth usage to assure existing calls receive high quality.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: Can I reuse one of my 2500 routers as a gatekeeper?

A: Yes, although you'll want to upgrade your IOS and memory. Most versions of IOS that support H.323 and MCM functions (such as gatekeeper) require 16MB of DRAM memory as well as 16MB of flash memory. Also note that (as with any IOS function) performance will be limited on the 2500 platform. For example, the 2500 will not support the high-performance gatekeeper. You'll probably only want to use a 2500 if you have a small voice and video implementation.

Q: What percentage of my WAN bandwidth should I allow to be consumed by H.323 calls?

A: This is a difficult question as every network is very different. If you mostly have applications that are not time-sensitive (e-mail, data replication, backups, and so forth) you can dedicate a large portion of your bandwidth to H.323 traffic (75 percent, for example). If you have mostly time-sensitive applications (SAP, telnet, database queries, and so on) you should only dedicate a modest amount of bandwidth (30 to 40 percent, for example). The percent of utilization currently on your WAN circuits will also affect how much bandwidth H.323 can use. If in doubt, only allow a small amount to be used by H.323, then if the network (traditional data and H.323) are running well, you can slowly increase the amount available to H.323.

For voice calls, remember there are many types of CODECs (analog to digital converters). Each one offers a different trade-off between voice quality and bandwidth required. You can experiment with different types (software selectable in Cisco Gateways) to determine the minimum bandwidth required for acceptable quality calls.

Q: Can I use the same gatekeeper for my voice and video calls?

A: Yes, any platform that will support the MCM gatekeeper will administer both of these calls. You will probably have different policies for each type of call (admission policies, dialing plans, and so on) though this is simply an administrative task; even the low-end gatekeepers will not have a problem with this. The only concern is that multiple types of calls (voice and video) typically lead to a greater number of calls, which require more resources.

Chapter 6

DSPs Explained

Solutions in this chapter:

- DSP Provisioning
- Conferencing and Transcoding
- Catalyst 4000 Modules
- Catalyst 6000 Modules
- NM-HDV Modules
- Sample Design Scenarios

- ☑ Summary
- ☑ Solutions Fast Track
- ☑ Frequently Asked Questions

Introduction

Digital Signaling Processor (DSP) provisioning is something that people have the tendency to neglect and is consequently added at the end of a project. By the time you finish reading this chapter, we hope you will have a better understanding of what a DSP is, on which platforms DSP resources are provided, and some of the more common scenarios in which DSPs are used.

Armed with this information, it is then possible to find out the requirements of your AVVID network and make the correct decision when provisioning your DSPs. In this chapter, we will discuss the available DSP resource solutions. These solutions include modules for the Catalyst 4000 and 6000, and NM-HDV modules, which are available in the 2600/3600/VG200 Series routers/gateways. This will be a good place to start when you are looking to understand how DSP resources will be involved in the planning and layout of your Voice over IP (VoIP) solution. We will discuss what DSPs are, what is provisioning, why they are necessary, and why planning and designing their usage in the overall AVVID design is important.

DSP Provisioning

A DSP is used to translate voice and fax signals into VoIP data streams. The number of conversions a DSP can perform is based on which CODEC complexity is being used. Cisco supports medium and high CODEC complexity. The following CODECs are medium complexity: G.711 (a-law and μ-law), G.726, G.729a, G.729ab, and Fax-relay. High-complexity CODECs are G.728, G.723, G.729, G.729b, and Fax-relay (medium-complexity CODECs can be run in high-complexity mode, but with fewer available channels). These coder-decoder compression algorithms convert the voice signals to packets ranging in size from 64K to 5.3K. The level of complexity, which is affected by the algorithm used by the compression CODEC, determines the number of calls a DSP can process. By using medium complexity, each DSP can process four calls, whereas with high complexity, only two calls can be processed per DSP. As you can see, the compression method used will affect the bit rate and quality of the call. Meanwhile, the quality of the voice conversation is benchmarked against the MOS (Mean Opinion Score) chart, shown in Table 6.1. This MOS rating is based on listeners judging the quality of a voice call.

Table 6.1 MOS Rating

Compression Method	Bit Rate (Kbps)	MOS Score
G.711 PCM	64	4.1
G.726 ADPCM	32	3.85
G.728 LD-CELP	16	3.61
G.729 CS-ACELP	8	3.92
G.729a CS-ACELP	8	3.7
G.723.1 MP-MLQ	6.3	3.9
G.723.1 ACELP	5.3	3.65

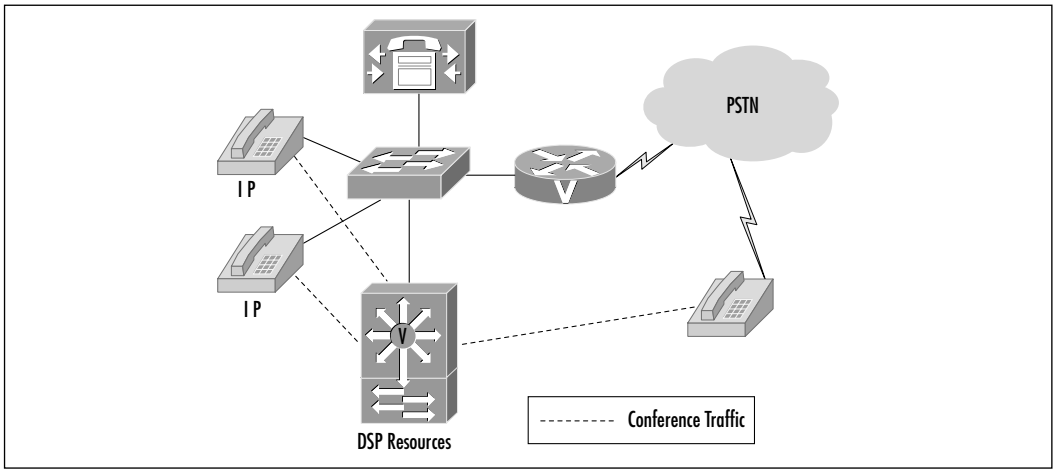
Cisco voice DSP modules are 72-pin Single In-line Memory Modules (SIMMs) from Texas Instruments (TI). Cisco uses TI DSPs, C542 and C549. The C542 DSP is supported on the AS5300 Voice Feature Card (VFC), 2600/3600/VG-200 Voice Network Modules: NM-1V and NM-2V, and MC3810 Voice Compression Module (VCM). The Cisco AS5800, AS5300 double-density VFC, 2600/3600/VG-200 High Density Voice Network Module (NM-HDV), 7200/7500 Digital T1/E1 High Capacity Voice port adapter (PA), 175x Series router, and MC3810 High-density Compression Module (HCM) utilizes the C549 DSP module. DSP SIMM modules are also in Catalyst 4000 and 6000 blades, WS-X4604-GWY and WS-X6608-T1/E1, respectively. This chapter will focus on the DSP modules in the Catalyst 4000 and 6000, and NM-HDV modules used as DSP resources or a DSP farm in CallManager deployments. The Catalyst 4000 and 6000 switches serve as DSP farms. A DSP farm is a pool or group of DSP SIMMs located in Cisco switch and router modules. Prior to CallManager 3.1, DSP resources were provisioned to individual Cisco CallManagers—in other words, the DSP resources could not be shared between servers. CallManager 3.1 allows resource sharing between servers in the cluster. DSP provisioning is the process of allocating DSP resources for Cisco CallManager clusters to provide hardware conferencing and transcoding functions. CallManager also supports software-based conferencing and transcoding. Just as with planning and designing your AVVID network in regards to what switches, routers, and gateways should be used, you should consider early on what DSP resources will be needed for a successful AVVID installation. To help in the planning of DSP resources and provisioning, you must consider all the applications and users that will be communicating with G.711 applications, such as voice messaging and interactive voice response (IVR) packages and IP WAN conferencing. We will discuss next what conferencing and transcoding

means to your AVVID network, as well as some possible scenarios and hardware solutions.

Conferencing and Transcoding

Conferencing is allowing multiple participants to join a common call. Cisco CallManager supports two types of conferencing: ad-hoc and meet-me. In ad-hoc conferencing, the originating caller controls the conference, and determines who will be on the call. The participants may even continue the call after the originating caller hangs up. A meet-me conference, meanwhile, allows participants to join into a conference call by calling into an assigned number out of a pool of directory numbers. More participants can continue to join the conference call until the maximum number allowed is reached. DSP resources support both types of conferencing, and the Cisco CallManager uses DSP resources to provide conferencing services, as shown in Figure 6.1. In this scenario, an IP phone caller joins another IP phone and an outside or PSTN initiated caller in a three-way conference call. This is an example of an ad-hoc conference. The Catalyst DSP resources are one way a Cisco CallManager is able to provide a conference bridge. A four-way G.711 conference call would utilize four DSP resources, one for each participant to stream into a single call.

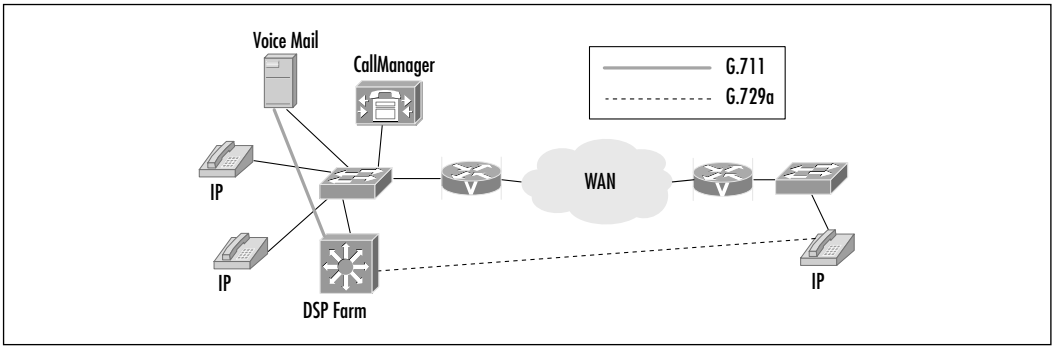
Figure 6.1 Conferencing



Cisco CallManager supports up to 48 software-based conference audio streams, which equates to 16 conferences with 3 participants (48/3). A non-CallManager server, on the other hand, supports up to 128 audio streams. Software conferencing

is based on G.711, whereas hardware-based solutions support G.711, G.729a, and G.723. The newer Cisco IP phones 7900 Series supports G.711 and G.729a, but the older style supports G.711 and G.723. Enterprises with large IP telephony environments must rely on hardware-based conferencing and transcoding solutions such as the Cisco Catalyst 4000 and 6000 switches. A hardware solution may be required based on your user demands, since a CallManager can only support up to 48 audio streams. This calculates out to 16 conferences maximum based on a 3-participant conference. Transcoding is the process of converting IP packets of voice streams between a low bit-rate (LBR) CODEC to and from a G.711 CODEC. A LBR is a CODEC such as G.729a or G.723. An example of a need for transcoding is when a user across the IP WAN wants to access a voice mail system which only supports G.711, and CallManager is configured to initiate remote IP calls using a G.729a CODEC, as shown in Figure 6.2. In this scenario, transcoding must be performed to convert the G.729a voice stream to G.711 in order to communicate with the voice mail system.

Figure 6.2 Transcoding



Deciding which DSP solution to implement is generally based on the size of the deployment; for instance, smaller size companies may want to consider deploying a software-based solution. The Cisco CallManager is able to handle 24 MTP sessions, while a separate Windows server can support 48 sessions. It's recommended to run this service on a separate server since performance will be affected on the CallManager for call processing. For medium size organizations, a hardware solution such as the Catalyst 4000 would be more feasible.

The larger enterprise should consider the Catalyst 6000 solution with the 8-port T1/E1 Voice and Service module. The Catalyst 6000 module is a highly scalable solution. We will next discuss the limitations of each of these modules.

Catalyst 4000 Modules

The Catalyst 4000 currently offers a mid-range hardware-based conferencing and transcoding solution. The Catalyst 4000 Access Gateway Module (AGM) provides the following services to the Catalyst 4000 switch: IP WAN routing, VoIP, and IP telephony. The Catalyst 4000 AGM supports voice interface cards (VICs) and WAN interface cards (WICs) from the 1600/1700/2600/3600 Series routers. The AGM supports the following ports and slots:

- **Two VIC/WIC slots** (VWICs, VICs, and WICs)
- **One dedicated VIC slot** (VWICs and VICs)
- **FlexSlot High Density Analog** (8-port RJ-21 FXS module)
- **Four DSP SIMM slots**
- **64 or 128MB memory SIMM slot** (Integrated Service Adapter)

VoIP gateway mode requires DSP resources to convert voice calls into data packets. The Cisco IOS IP/DSP Plus feature set is another requirement to allow VoIP gateway capabilities. The AGM supports the following interfaces as a VoIP gateway:

- T1 and E1 ISDN Primary Rate Interface (PRI)
- T1 Channel Associated Signaling (CAS)
- Foreign Exchange Office (FXO)
- Foreign Exchange Station (FXS)
- Ear & Mouth (E&M)

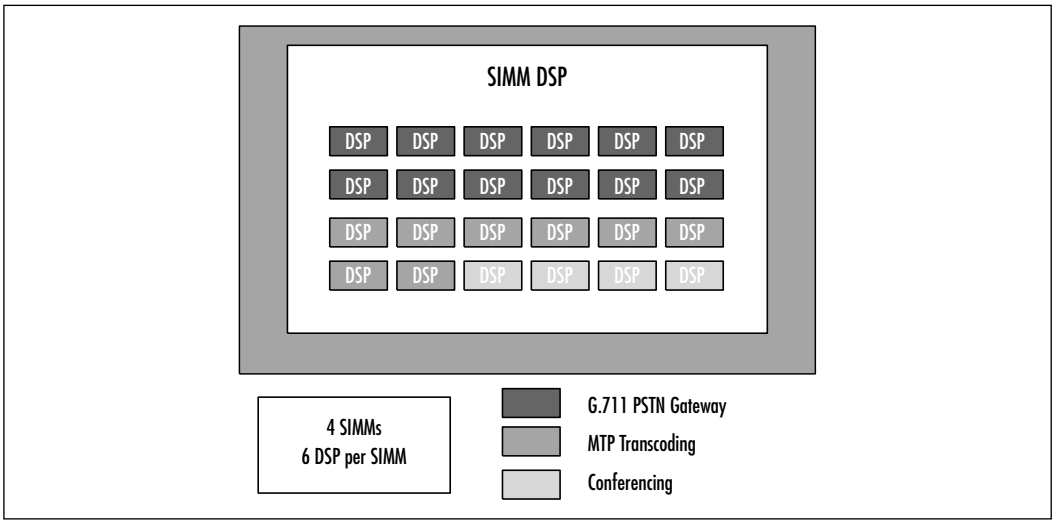
The AGM also supports functioning as a DSP farm for Cisco CallManager. In turn, Cisco CallManager can be configured to provide hardware-based meet-me and ad-hoc conferencing features. The Catalyst 4000 gateway module has four DSP SIMMs with each SIMM having six DSPs for a total of 24 DSP resources. Table 6.2 summarizes the Catalyst 4000 AGM capabilities.

Table 6.2 Catalyst 4000 AGM DSP Resources

Function	Capability
PSTN gateway	96 channels of G.711 voice
Conferencing	24 channels of G.711 conferencing (4 conferences x 6 through 8 x 3)
MTP transcoding	16 channels of LBR to G.711

Figure 6.3 depicts how the DSP resources are provisioning within the Catalyst 4000 after it is configured for gateway mode.

Figure 6.3 Gateway Mode DSP Resources



The Catalyst 4000 DSPs can only support G.711 conferencing sessions. This is not to say that there can be a conference session on a Catalyst 4000 with only G.711 participants, but that transcoding DSP resources must be involved to convert those participants to the G.711 CODEC for the conference. The Catalyst 4000 AGM module supports up to 16 transcoding sessions per module via hardware and would handle up to 104 channels in software.

Configuring & Implementing…

Configuring the Catalyst 4000 AGM

In order to provision DSP resources for conferencing and transcoding, you must configure the AGM for IP telephony gateway mode. You must first set up the Access Gateway module in Cisco CallManager as an H.323 gateway with H.225 device protocol. For more details on configuring Cisco gateways in CallManager, see www.cisco.com/univercd/cc/td/doc/product/voice/c_callmg/3_1/sys_ad/adm_sys/ccmcfg/b06gtway.htm.

Next, you must set up the switch to communicate with the CallManager and set the IP precedence. The following is an example of the necessary steps:

```
Cat4K_AGM(config)# voicecard sccp manager 192.168.10.250
                   port 2000
Cat4K_AGM(config)# voicecard sccp local g0/0
Cat4K_AGM(config)# voicecard conference ip-precedence
```

Configuring the AGM as a Conference Bridge requires some configuration on the module before it is configured in CallManager. You will need the MAC address in order to complete the configuration in CallManager. This can be obtained by entering the following command:

```
Cat4K_AGM# show voicecard conference
Conferencing:(mac address 00e0.dbef.4863.93e9) Disabled
```

Once you have the MAC address of the Catalyst 4000 AGM, complete the steps of setting up a conference bridge in the CallManager Administration program. For more detailed steps for this process see www.cisco.com/univercd/cc/td/doc/product/voice/c_callmg/3_1/sys_ad/adm_sys/ccmcfg/b04cnbrg.htm.

Keep in mind, the Catalyst 4000 must be in gateway mode. This can be accomplished by entering either or both of the following commands:

```
Cat4K_AGM(config)#voicecard conference ;Enable conferencing
Cat4kagm(config) #voicecard transcode ;Enable Transcoding
```

Catalyst 6000 Modules

The Cisco WS-6608-T1/E1 module for Catalyst 6000 offers similar functionality as the Catalyst 4000 AGM. The Catalyst 4000 AGM is intended to be implemented at the remote or branch office environment and the Catalyst 6000 Voice T1/E1 and Services module is targeted at the central office or headquarters level environments. The Catalyst Voice T1/E1and Services module provides Digital T1 or E1 PSTN and PBX gateway services, transcoding, and conference bridging. The Voice T1/E1 and Services module has eight T1 or E1 ports which support common channel signaling (CCS) or ISDN PRI signaling. Each port can be configured as a PSTN/PBX gateway, MTP transcoder, or a conference bridge. Table 6.3 summarizes the DSP resource capabilities of the Catalyst 6000 Voice

T1/E1 and Services module. Note that some solutions will work with either T1or E1, or in some cases work only with T1 or E1.

Table 6.3 Catalyst 6000 DSP Resources

Function	Capability
PSTN gateway	WS-6608-T1 module: ■ 24 calls per DS1 port ■ 192 calls per module WS-6608-E1: ■ 30 calls per DS1 port ■ 240 calls per module
Conferencing	G.711 or G.723: ■ 32 conferencing participants per physical port ■ Maximum conference size of six participants ■ 256 conference participants per module G.729: ■ 24 conferencing participants per physical port ■ Maximum conference size of six participants ■ 192 conference participants per module
MTP transcoding	G.723 to G.711: ■ 31 MTP transcoding sessions per physical port ■ 248 sessions per module G.729 to G.711: ■ 24 MTP transcoding sessions per physical port ■ 192 sessions per module

The 8-port T1/E1 Voice and Services module performs 24 transcoding sessions per port when translating from G.729 to G.711. It does 31 sessions per port for G.723 to G.711. So the 8-port module can scale to 192 or 248 sessions per module depending on the LBR CODEC be utilized. Besides the higher density, another difference between the Catalyst 6000 solution and the 4000 is that the Catalyst 6000 is able to perform a mix of transcoding and conferencing within the same DSP. You may wonder why this is important—as an example, if we have a four-user conference call with two users crossing an IP WAN configured for a LBR, the Catalyst 4000 would use one DSP for each user crossing the IP WAN for converting their call to G.711, and four DSPs to convert the G.711 streams into a time division multiplexing (TDM) stream to be summarized, and finally four DSPs to mix the callers together, for a total of ten DSP resources. However, the Catalyst 6000 will use four DSP resources to transcode and conference all the callers into a single call.

Each port on WS-6608-T1/E1 is configured with an IP address in order to configure it as a PSTN gateway, conferencing, or transcoding resource in Cisco CallManager. The address can be a static assigned address, or via DHCP. In addition to the static IP address, a TFTP server must be configured since the configuration is downloaded from a TFTP server.

NOTE

When designing conferencing services utilizing the Catalyst 6000, keep in mind that the conferencing service does not span ports. This means that any given conference cannot use resources from another port. Therefore, a conference cannot support more participants than any given port will.

Configuring and Implementing…

Configuring the Catalyst 6000 Voice T1 and Services Module

To illustrate the difference between the Catalyst 4000 module and the Catalyst 6000, we will describe the steps to configure the Voice T1/E1 module, which is done on a port-by-port basis. The port can be configured for PSTN connectivity, transcoding, or conferencing. The transcoding port shows up as "MTP" while the conference port shows up as "Conf Bridge."

First, let's look at a sample configuration of a Catalyst 6000 with an 8-port T1 Voice and Service module. The following command shows the status of the module located in slot 7 of the Catalyst 6000.

```
Cat6K> (enable) show port 7

Port  Name              Status      Vlan     Duplex Speed  Type
----  ----------------- ----------  -------- ------ -----  -----
 7/1                    notconnect  1        full   1.544  T1
 7/2                    notconnect  1        full   1.544  T1
 7/3                    connected   1        full   1.544  T1
 7/4                    connected   1        full   1.544  T1
```

Continued

```
 7/5                   connected   1         full 1.544 T1
 7/6                   connected   1         full 1.544 T1
 7/7                   enabled     1         full    - Conf Bridge
 7/8                   enabled     1         full    - MTP

Port   DHCP    MAC-Address        IP-Address      Subnet-Mask
------ ------- ------------------ --------------- ------------
 7/1   enable  00-10-7b-00-12-08 10.6.15.165      255.255.255.0
 7/2   enable  00-10-7b-00-12-09 10.6.15.166      255.255.255.0
 7/3   enable  00-10-7b-00-12-0a 10.6.15.167      255.255.255.0
 7/4   enable  00-10-7b-00-12-0b 10.6.15.168      255.255.255.0
 7/5   enable  00-10-7b-00-12-0c 10.6.15.169      255.255.255.0
 7/6   enable  00-10-7b-00-12-0d 10.6.15.170      255.255.255.0
 7/7   enable  00-10-7b-00-12-0e 10.6.15.171      255.255.255.0
 7/8   enable  00-10-7b-00-12-0f 10.6.15.172      255.255.255.0

Port   CallManager(s)  DHCP-Server    TFTP-Server    Gateway
------ ---------------- -------------- -------------- ---------
 7/1   10.6.15.155       10.6.15.155    10.6.15.155        -
 7/2   10.6.15.155       10.6.15.155    10.6.15.155        -
 7/3   10.6.15.155       10.6.15.155    10.6.15.155        -
 7/4   10.6.15.155       10.6.15.155    10.6.15.155        -
 7/5   10.6.15.155       10.6.15.155    10.6.15.155        -
 7/6   10.6.15.155       10.6.15.155    10.6.15.155        -
 7/7   10.6.15.155       10.6.15.155    10.6.15.155        -
 7/8   10.6.15.155       10.6.15.155    10.6.15.155        -
.
.
.
```

Continued

```
Port      CallManagerState   DSP-Type
--------- ------------------ --------
 7/1      registered           C549
 7/2      registered           C549
 7/3      registered           C549
 7/4      registered           C549
 7/5      registered           C549
 7/6      registered           C549
 7/7      registered           C549
 7/8      registered           C549


Cat6K> (enable)
```

Similarly, the Catalyst 4000 the 6000 must be set up in CallManager Administration as a Conference Bridge.

Again, for the detailed steps refer to the following link: www.cisco.com/univered/cc/td/doc/product/voice/c_callmg/3_1/sys_ad/adm_sys/ccmcfg/b04cnbrg.htm.

To verify the configuration of the Catalyst 6000 and show active calls enter the following command:

```
Cat6K> show port voice active


Port   Type         Total   Conference-ID/   Party-ID   IP-Address
                            Transcoding-ID
------ ------------ ------- ---------------- ---------- -----------
 3/1   call         1       -                -          199.22.25.254
 3/2   call         1       -                -          178.205.25.54
 4/5   call         3       -                -         165.34.234.111
                                                          172.32.34.12
                                                         198.96.23.111
 3/8   conferencing 2       1                1        200.200.200.241
                                               2          173.23.13.42
                                               3         198.97.123.98
                                               5          182.34.54.26
```

Continued

```
3/2  call          1      -                -        178.205.25.54
3/8  transcoding    1      1                1      200.200.200.241
                                            2         183.32.43.3
```

NM-HDV Modules

One of the benefits of the NM-HDV module is toll bypass and PBX leased line replacement, which reduces toll charges by local and long distance carriers and eliminates monthly leased lines charges for PBX interconnections. Another is acting as a PSTN gateway for Cisco CallManager and IP phones. As an enterprise grows, this is a great way for VoIP to be seen as a low-cost solution to the extraordinary expense of long distance toll charges.

The High Density Voice Network Module (NM-HDV) family supports the Multi-flex Voice/WAN (VWIC) and VIC cards. The NM-HDV module is also known as the Digital T1/E1 Packet Voice Trunk Network Module. The platforms supported by the NM-HDV are VG-200, 2600, and 3600 series routers. The NM-HDV supports connections to both private branch exchanges (PBXs) and public switched telephone networks (PSTN). A single module can scale up to 60 voice channels. These 60 voice channels are based on medium-complexity voice compression CODECs. Each NM-HDV module contains at least one 12-channel Packet Voice DSP Module (PVDM-12) with a maximum of five PVDM-12s possible. The PVDM-12 has three TI 549 DSPs, and supports 12 medium-complexity and 6 high-complexity voice calls. Up to 6 packet voice trunk modules can be configured in a Cisco 3660 router supporting 288 voice channels. Table 6.4 lists all the available configuration options for the Digital T1/E1 Packet Voice Trunk modules.

Table 6.4 Digital T1/E1 Packet Voice Trunk Module Family

Part Number	Description
NM-HDV-1E1-12	High Density Voice Network Module, with 1 VWIC-1MFT-E1 and 1 PVDM-12
NM-HDV-1E1-30	High Density Voice Network Module, with 1 VWIC-1MFT-E1 and 3 PVDM-12
NM-HDV-1E1-30E	High Density Voice Network Module, with 1 VWIC-1MFT-E1 and 5 PVDM-12
NM-HDV-2E1-60	High Density Voice Network Module, with 1 VWIC-2MFT-E1-DI and 5 PVDM-12

Continued

Table 6.4 Continued

Part Number	Description
NM-HDV-1T1-12	High Density Voice Network Module, with 1 VWIC-1MFT-T1 and 1 PVDM-12
NM-HDV-1T1-24	High Density Voice Network Module, with 1 VWIC-1MFT-T1 and 2 PVDM-12
NM-HDV-1T1-24E	High Density Voice Network Module, with 1 VWIC-1MFT-T1 and 4 PVDM-12
NM-HDV-2T1-48	High Density Voice Network Module, with 1 VWIC-2MFT-T1-DI and 4 PVDM-12
NM-HDV	High Density Voice Network Module (no VWIC, no PVDM)
PVDM-12	12-Channel Packet Voice DSP Module

When designing an AVVID network, it's important to understand what versions of software or hardware components are required for your desired features. Table 6.5 lists the Cisco hardware platforms and their IOS versions, which support the various NM-HDV modules.

Table 6.5 NM-HDV Platform Support

IOS Support	VG200	2600	3620, 3640	3660
NM-HDV-1E1-30	12.1(5)XM1	12.0(7)XK, 12.1(2)T	12.0(7)XK, 12.1(2)T	12.0(7)XK, 12.1(2)T
NM-HDV-1E1-30E	12.1(5)XM	12.0(7)XK, 12.1(2)T	12.0(7)XK, 12.1(2)T	12.0(7)XK, 12.1(2)T
NM-HDV-2E1-60	12.1(5)XM1	12.0(7)XK, 12.1(2)T	12.0(7)XK, 12.1(2)T	12.0(7)XK, 12.1(2)T
NM-HDV-1T1-24	12.1(3)T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T
NM-HDV-1T1-24E	12.1(3)T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T
NM-HDV-2T1-48	12.1(3)T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T

Continued

Table 6.5 Continued

IOS Support	VG200	2600	3620, 3640	3660
NM-HDV	12.1(3)T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T
PVDM-12	12.1(3)T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T	12.0(5)XK, 12.0(7)T, 12.1, 12.1T

The NM-HDV DSPs are utilized for incoming PSTN calls being converted into G.711 or G.729 VoIP calls. These are the calls coming in from the T1 or E1 circuit on the NM-HDV module. The DSPs on the NM-HDV can support up to four voice calls per DSP.

Sample Design Scenarios

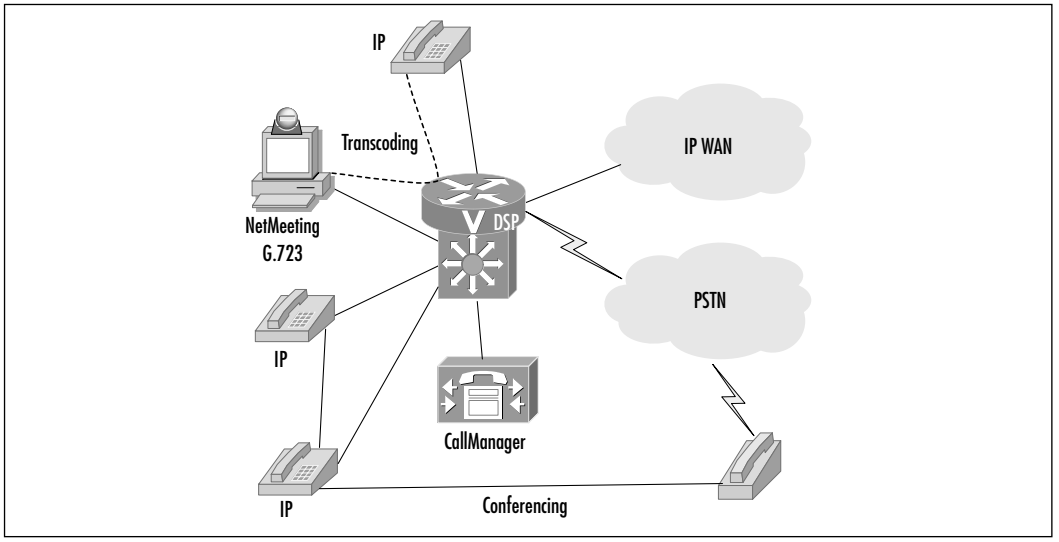
We will discuss two design scenarios and the requirements they are trying to meet. The first scenario is a medium sized branch office with no legacy PBX equipment. The second scenario is a large enterprise campus infrastructure with a legacy PBX and CallManager. The following examples should give you a better understanding of what type of solution may be applicable to your environment. These solutions are geared towards implementing a hardware-based DSP solution. If you are in a small environment, using the features and capabilities of your CallManager may work for your situation.

Branch Office

If you are in a central branch office and your needs are to have an integrated and simplified solution, the Catalyst 4000 switch with the AGM is an excellent choice. The Catalyst 4000 is a recommended solution for a branch office with connections to the PSTN and possibly an IP WAN circuit to another office. For example, a branch office that has 175 employees with 140 IP phones and PCs in turn requires the ability to perform multiparticipant conferencing and transcoding. The transcoding is a requirement because they have some users running Microsoft NetMeeting with G.723 CODEC performing audio conferencing with IP phones. The Catalyst 4006 with a CallManager can provide all the needs for this branch office environment. The AGM running in IP telephony gateway mode will provide the DSP resources to allow conferencing and transcoding. Figure 6.4 illustrates how the Catalyst 4000 is the backbone of the IP telephony solution supplying PSTN,

WAN, LAN connectivity, and DSP farm resources in this single box solution. 24 conference participants and 16 transcoding sessions can be supported per AGM module. When planning and designing your AVVID network, which includes calculating your DSP resources, you need to put adequate effort in the analysis of how the telephony network will be utilized and where the need for transcoding lies. Accurate estimates of the number of conferencing sessions need to be performed. If undersized, you will not be providing the level of service necessary to have satisfied users. If oversized, you may not be spending your investment dollars wisely. This environment is conducive to the Catalyst 4000 AGM, since conferencing should be completed with all participants running a G.711 CODEC and not requiring any transcoding, which will use more DSP resources.

Figure 6.4 Branch Office

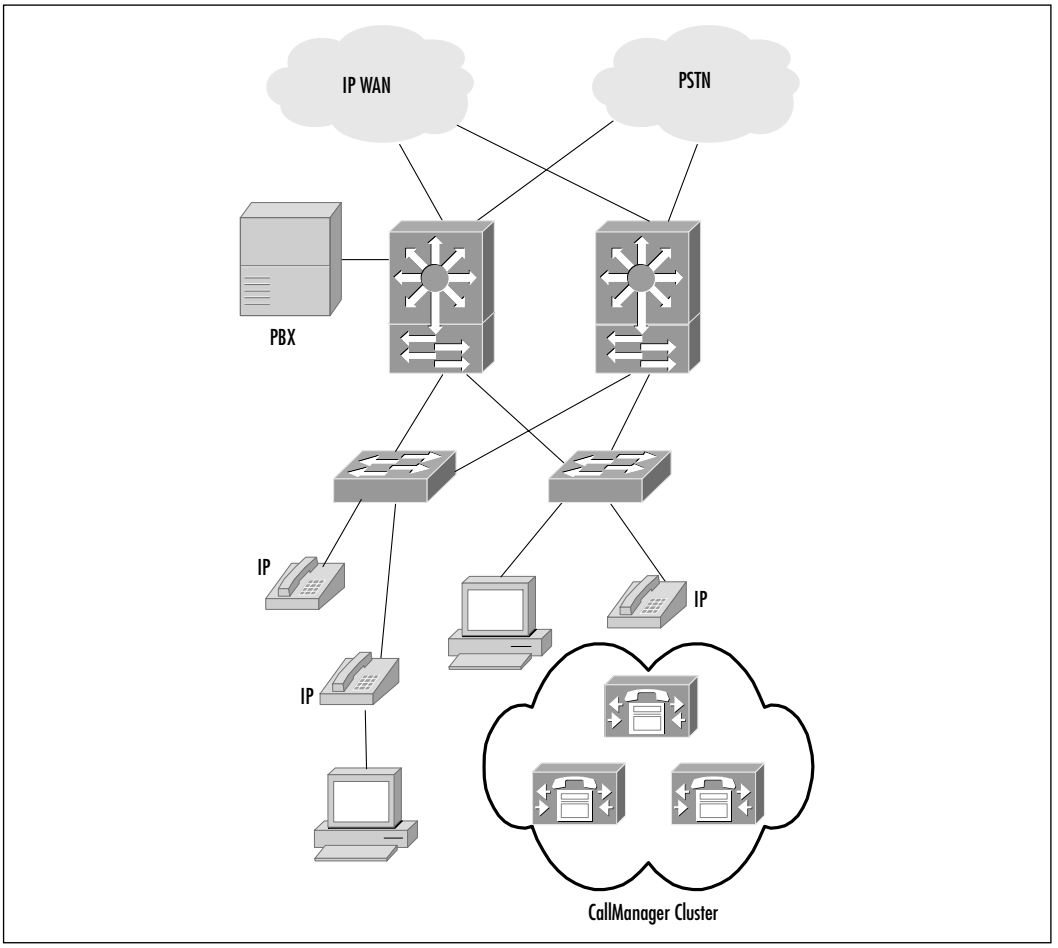


Enterprise Campus

If your environment is a larger campus infrastructure with hundreds to thousands of users, the Catalyst 6000 is geared toward this size network. The Catalyst 6000 Series can scale to hundreds of users per switch and provides the level of DSP resources required for this size of deployment. The Catalyst 6000 is the backbone of the IP telephony network, and supports the Cisco IP phones for power and LAN connectivity. The Catalyst 6000 with the 8-port T1/E1 Voice and Services module can support 256 conference participants. Depending on your company's

call traffic patterns, a single T1/E1 module could satisfy your needs. It communicates with the Cisco CallManager to allocate DSP resources for conference bridges and transcoding. Figure 6.5 illustrates a large campus infrastructure with a CallManager cluster and multiple Catalyst 6000 switches with T1/E1 modules acting as a DSP resource farm, in addition to providing PSTN and IP WAN connectivity. The Catalyst 6000 has the scalability and performance to satisfy the larger environment. In this CallManager cluster network, the DSP resources can be provisioned and shared between the servers.

Figure 6.5 Enterprise Campus



Summary

As explained in this chapter, the more complex and integrated your AVVID network becomes, the more important your DSP provisioning decisions become, and we feel an integral part of the planning process. Knowing ahead of time what you expect your network capabilities to be for conferencing, transcoding, and support will lend itself directly to the direction your DSP needs are sure to follow. These decisions should largely be based on the size and scope of the deployment. For smaller-sized companies, you may wish to consider a software-based solution such as Cisco's CallManager on a Windows server. Medium-sized deployments could be handled with a Cisco Catalyst 4000 switch, while the large enterprise would obviously best be served by a Cisco Catalyst 6000 with the 8-port T1 Voice and Service module. Clearly, whatever the size and scope of the deployment, making a few key DSP Provisioning decisions in the planning or early stage design of the network will save many headaches and sleepless nights as the project grows.

Solutions Fast Track

DSP Provisioning

- ☑ The Cisco DSP module is a Texas Instruments model C542 and C549 72-pin SIMM. These DSPs work with two levels of CODEC complexity: medium and high.
- ☑ The medium-complexity CODECs that work with the Cisco DSP are G.711 (a-law and μ-law), G.726, G.729a, G.729ab, and Fax-relay. The high-complexity CODECs include the G.728, G.723, G.729, G.729b, and Fax-relay.
- ☑ The DSP resources are used for conference bridging and transcoding.

Conferencing and Transcoding

- ☑ Conferencing is the process of joining multiple callers into a single multiway call. The two types of multiparticipant voice calls supported by the Cisco CallManager are ad-hoc and meet-me.

- ☑ DSP resources are used in the conference bridge scenario to convert VoIP calls into TDM streams and sum them into a single call.
- ☑ Transcoding is the process of converting IP packets of voice streams between a low bit-rate (LBR) CODEC to G.711. Transcoding functions can be done by converting G.723 and G.729 CODECs to G.711.
- ☑ Conferencing and transcoding is performed either by hardware or software. The software version is performed on a Cisco CallManager server, while the hardware solutions are the Catalyst 4000 AGM module, Catalyst 6000 8-port T1/E1Voice and Services module, and NM-HDV module.

Catalyst 4000 Modules

- ☑ The Catalyst 4000 Access Gateway Module (AGM) provides voice network services to the Catalyst 4000 switch, VoIP IP WAN routing, and an IP telephony mode for use with a voice gateway. The Catalyst 4000 AGM supports voice interface cards (VICs) and WAN interface cards (WICs) from the 1600/1700/2600/3600 Series routers.

Catalyst 6000 Modules

- ☑ The Catalyst 6000 switch module features an 8-port Voice T1/E1 and Services module, WS-X6608-E1 or WS-X6608-T1.
- ☑ The Voice T1/E1 module supports T1/E1 CCS signaling, ISDN PRI network, and user-side signaling. Similar to the AGM module for the Catalyst 4000, the Voice T1/E1 can be provisioned for conferencing and transcoding. The Voice T1/E1 can do mixed CODEC conferencing, whereas the AGM only does G.711 conferencing with individual DSP resources.

NM-HDV Modules

- ☑ The biggest benefit of this module is PBX leased line replacement and toll bypass, meaning that a company's long distance expenses can all but be eliminated. Platform support includes VG200, 2600, 3600, and Catalyst AGM E1 Models (medium complexity involving NM-HDV-1E1-12, NM-HDV-1E1-30, and NM-HDV-2E1-60). With E1 Models

(high complexity M-HDV-1E1-30E), or T1 Models, and medium complexity (NM-HDV-1T1-12, NM-HDV-1T1-24, and NM-HDV-2T1-48) supported, it will also support T1 Models (high complexity NM-HDV-1T1-24E).

Sample Design Scenarios

- ☑ When designing your DSP provisioning, you must take into account the number of users, the type of applications using different CODEC, and the overall IP telephony design to determine which solution best fits your needs, whether it's using the CallManager itself or one of the Catalyst switches.
- ☑ The branch office environment is an excellent candidate for the Catalyst 4000 switch with an Access Gateway module (AGM). This solution can provide 10/100/1000 Ethernet switching with inline power for IP phones, PSTN connectivity, IP routing, and also serve as a DSP resource. The DSP resources provide conferencing and transcoding services for your user population.
- ☑ The enterprise campus has higher scalability requirements than the branch office. With this in mind, you should consider the Catalyst 6000 with the 8-port T1/E1 Voice and Service module as a good fit for the needs of this environment.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: Without a Cat 4006 or 6500, where do I get DSP resources for conference calls?

A: DSPs are usually hardware-based, generally on the Catalyst 4000 and 6000 Series switches. The conference bridge can be supported on a small scale on the CallManager itself, but that is not recommended since it can impact the performance of the CallManager.

Q: How many calls can be made with a NM-HDV-2E1-60?

A: The NM-HDV-2E1-60 with 15 DSPs and two E1 circuits can support 60 medium-complexity calls. This is based on the 15 DSPs each supporting four calls.

Q: When do I need the DSP option for the Access Gateway Module?

A: The DSP option is required when voice functionality is enabled. This includes any voice gateway functions or voice network services.

Q: What feature set of IOS is required to use DSPs in the Catalyst 4000 AGM?

A: The Cisco IOS IP/Firewall/DSP Plus feature set is required.

Chapter 7

AVVID Applications

Solutions in this chapter:

- Creating Customer Contact Solutions
- Providing Voice Recording Options
- Call Accounting, Billing, and Network Management Solutions
- Designing Voice and Unified Messaging Solutions
- Understanding Other Voice Applications

- ☑ Summary
- ☑ Solutions Fast Track
- ☑ Frequently Asked Questions

Introduction

In this chapter, we hope to give you insight into some of the advantages of the provided applications by utilizing a converged solution, as well as present you with an understanding of some of the design considerations associated with each of the different applications, including Interactive Voice Response (IVR), Web Attendant, Administrative Reporting Tool (ART), and Voice Recording solutions.

This range of applications is one of the major factors that differentiate an Internet Protocol (IP) Telephony solution from the traditional solutions of the past. Once you bought a traditional solution from a specific vendor, you were pretty much tied up in terms of who, what, and where you could buy any of your future applications. These decisions also tied down the number of supported applications for the specific platforms you would normally work with. Now there is a converged solution that is able to support many of the standards in place today, as well as providing you the capability to support future standards once ratified. Currently these solutions are more software-based than hardware-centric, allowing many different applications to be integrated by using standards-based interfaces. No longer are you tied to a single vendor to provide a solution; you now have the ability to choose which vendor provides you with the most suitable answer to your specific requirements, and who is able to offer you the best future scalability.

Within the framework of AVVID, and more specifically, the Cisco CallManager, is the ability to provide application integration using either the Telephony Application Programming Interface (TAPI) and the Java Telephony Programming Interface (JTAPI), as well as several other supported standards-based solutions. There are several Cisco applications that can use these APIs, such as the Cisco IP SoftPhone (which uses TAPI) and the Cisco IP-IVR (which uses JTAPI). With these standards in place, many other vendors' applications are able to integrate via these APIs. Some of these vendors may even use Cisco's own Skinny Protocol for support and management of Cisco applications and products. One of the most notable of these applications was a company called Active Voice, which provided a Voice Mail/Unified Messaging Solution called *Unity* that utilized TAPI (as well as Skinny) to provide integration into Cisco CallManager. Unity was such a well-constructed and integrated application that Cisco has since acquired the solution and it is now marketed under the name Cisco Unity (we will be discussing Unity later in this chapter). Using such an open solution allows you to feel like you are more in control of your fate since it allows you to make

decisions as to where you would like your converged solution to be in the future, rather than being dictated to by closed proprietary solutions.

NOTE

It is also possible to use Cisco's Skinny Protocol to provide integration into the CallManager solution.

Open standards is obviously something Cisco would like to utilize. Based on this, we will later in this chapter look into some applications that have been integrated into Cisco's IP telephony solutions and how open standards were used to help them incorporate and streamline these products within many companies' network infrastructure and planned growth. We will also be providing links to some of the vendors that provide solutions which either complement the converged solution, or, in some cases, even compete with products Cisco already offers. As mentioned previously, the fact that vendors can integrate via standards-based APIs provides you with the ability to choose the applications relevant to you, and that suit your business needs.

Creating Customer Contact Solutions

The call center market, or *contact center* market as it is currently referred to, is an area where much interest is being generated. The reason for this is that IP is able to bring much more to the table than traditional solutions ever could because of its widespread deployment and robust features. Customers can now define how they want to interact with a company, as opposed to previous solutions where organizations had certain channels customers could use.

With new world contact centers, channels such as voice, e-mail, fax, and Web collaboration are now all possible. In traditional solutions, data (e-mail and Web traffic as well), voice, and video are carried on separate infrastructures, involving the purchase, installation, and management of multiple networks. A voice solution in the majority of the cases has no view or understanding of the data network and vice versa, Video solutions were traditionally room-based, and had NO integration into any other business processes. In addition, there is no unified view of the user from a contact center perspective, requiring that business rules be configured in multiple places.

Using Cisco AVVID, it is possible to use IP as an enabler to let customers define how they want to do business with you and not vice versa. As the old saying goes, "The customer is king", and if the customer would like to send you an e-mail, as opposed to giving the call center a ring, they should still receive the same type of service regardless of the contact means. You need to be able to cater for this in a way that is both transparent to the customer yet provides a painless integration for the company.

Normally, the e-mail queue is serviced separately from the voice queue, so you either have a dedicated agent or multiple agents who deal with e-mail solely, or when agents get a free moment, they look in the e-mail queue to see if anything has arrived in the contact center mailbox that needs servicing immediately. The problem here is that if you have a person dealing with the e-mail queue, chances are they are only skilled in a certain area, and, based on this, will only be able to service queries in that specific area; anything else will have to be passed on to a particular agent, who will service this once they have a chance. During busy times, this could take awhile, which means you may be dealing with five calls worth say $10,000 while there is a $100,000 order waiting in the e-mail queue. In an ideal world, agents need to be notified that there is a large order waiting in the e-mail queue, and that they should work on that as opposed to servicing smaller orders. In other words, work smarter, not harder. With some of the solutions available today, it is possible for an e-mail to receive the same treatment as a voice call. An example follows.

A person from Company XYZ sends an e-mail asking for a large order to be shipped to a site. At precisely the same time, a person dials into the contact center and needs to find out the balance of his account. In this scenario, we could do several things:

1. Company A could be set up to answer voice calls, so they can give the customer an answer. When this action is completed, and if time is available, they can have a look at the e-mail queue, and answer any queries there.
2. Company B may have an IVR in place. In this instance, the customer who wanted to receive a balance could be serviced via the IVR and his balance given to him without any intervention from the agent. However, the agent still needs to go to the e-mail queue and service it.

The most ideal situation is that we receive both the voice calls and e-mail simultaneously; the e-mail is scanned, based on predefined rules, so you can see who it is from, make a decision that this is from one of our major customers and

answer the e-mail immediately. Because the agent is dealing with an e-mail, a higher application will realize the agent is busy and not send a call through until the agent has finished with the e-mail. In the meantime, the customer who calls in can enter his account number and pin code into an IVR, choose to see their balance, and the IVR will prompt the balance back to the customer without any agent intervention, thereby greatly enhancing customer satisfaction by dealing with multiple contact channels. This action will also cut down on the time necessary to service incoming requests. In addition, you can provide the voice caller with the option to speak to an agent if all his needs are not serviced via the IVR.

Defining the Customer Contact Channels

To define which customer contact channels you need to use depends on many different criteria. However, some of the questions you should be asking are as follows:

- What channels are currently in use by the contact center today? How efficient are they? What do I need to modify to make them more efficient?
- By evaluating the type of customers you are servicing, and by adding additional contact channels (either voice, e-mail, fax, Web collaboration, and so on), would it make the agent and customer's interaction with each other easier and more efficient?
- By implementing an additional contact channel, would it increase customer satisfaction, bearing in mind that happy customers are more likely to recommend your service than unhappy ones?
- Can the additional cost of the hardware/software/integration customer contact channel be justified by a return-on-investment analysis?
- Agents are normally one of the most expensive parts of a contact center, especially monthly operating costs, so what can I do to better manage their time, allowing them to more effectively deal with customer queries?

Cisco IPCC

Cisco's vision for the IP contact center (IPCC) market includes the unification of voice and data to support how a customer wishes to contact a company anywhere, anytime, via any channel. Cisco introduced the Cisco IPCC not as a product, but as a solution that combines several packages for ease of administration and control.

The IPCC solution, as just mentioned, is made up of a number of key components. These are:

- Cisco CallManager software
- Cisco IP-IVR software
- Cisco Intelligent Contact Management software
- Agent Desktop Presentation
- IPCC Hardware requirements
- Underlying Infrastructure requirements
- Cisco E-mail Manager (Optional)
- Cisco Web Collaboration Solution (Optional)
- Cisco Unity (Optional)
- PBX Integration (Optional)

Cisco CallManager

The CallManager is, in simplistic terms, the traditional PBX component of the Cisco IPCC. It is currently either supplied as a software-only solution for supported hardware platforms, or as a hardware/software Cisco solution. The hardware requirements will be discussed in more detail later in the chapter.

The CallManager does not, as such, have any Automatic Call Distributor (ACD) functionality, but rather relies on other applications to provide these features in addition to other advanced functionality. The CallManager has the ability to provide the information for one phone to call another (not necessarily an IP phone). It maintains the dial plans, phone information (such as where a particular extension can or cannot call), and has the capability to manage phones as well as gateways, not to mention other capabilities. The CallManager is probably one of the most intelligent parts of an IP telephony solution being that it has all the knowledge of the network. A single CallManager can (with the correct hardware configuration) support up to 2,500 extensions, allowing the organization to expand seamlessly should the need arise. In a fully configured cluster, this number rises to 10,000. These figures change, however, when looking at the IPCC since the IPCC Agent Desktop (as well as some other applications) has additional weighting that needs to be taken into consideration. Weights for specific devices are discussed in Chapter 4, and should give you an idea of how many devices it's

possible to use on a specific CallManager in a non-IPCC solution. The following are some design considerations when looking at the maximum amount of agents supported in an IPCC:

- Up to a maximum of 200 agents per CallManager
- Maximum of 500 IPCC agents, regardless of the number of CallManagers
- Maximum of 2.77 calls per second per CallManager (This equates roughly to 10,000 Busy Hour Call Completions [BHCC].)

NOTE

At the time of this writing, the upper limit of supported agents was currently enforced by Cisco. However, in the short- to medium-term, these numbers will hopefully increase or even fall away so as to basically cater to the majority of IPCC installations. When implementing an IPCC, remember to ensure that all of the version numbers of the different components (CallManager, IP-IVR, ICM, and so on) are supported within the IPCC solution. For more details on the current supported versions, please refer to the following URL: www.cisco.com/warp/customer/78/sw_compatability_matrix.html.

A single CallManager can be used in a call center environment as well as being used in the traditional PBX type role. However, this should be done with caution. This type of environment, while probably cost efficient, could lead to disaster should there be any problems with the CallManager (for example, if someone inadvertently switches off the power). Since CallManager 3.*x*, it has been possible to have an IP phone register with multiple CallManagers. What this means is that even if the primary CallManager was switched off, the IP phone would still be able to operate via a secondary or even tertiary CallManager. With applications such as the IP-IVR, IP SoftPhone, IP-Integrated Contact Distribution (IP-ICD), and Personal Assistant, this was not possible. If the primary CallManager were switched off for example, this would mean that the application would not work. With the release of CallManager 3.1, this has changed. Within a CallManager cluster, there is now the option to use the CTI Manager to provide TAPI/JTAPI redundancy. This, in effect, means that if, for example, your Primary CallManager were to be switched off, your application (if it is using TAPI/JTAPI)

would be able to reconnect to another configured CallManager within a cluster. Features such as that just described allow the CallManager to provide redundancy and resiliency throughout a Cisco AVVID Network.

Cisco IP-IVR

The Cisco IP-IVR is, as the name implies, a Cisco Internet Protocol Interactive Voice Response (IP-IVR) unit. This provides companies with the ability to prompt an incoming call for some type of information, collect the information, and possibly do a database lookup, or pass on this information elsewhere. While this is important, the most valuable feature of the IP-IVR is that it is the queue point for IPCC. When there are no agents available to service a call, the IP-IVR will hold that call, as well as provide some type of music-on-hold until instructed by another application (possibly Intelligent Contact Manager [ICM]) to pass the call to another source. Have you ever given a call center a ring, and instead of speaking to a person, gotten one of these standard prompts:

- "Good Day and welcome to XYZ Corporation. We apologize, but currently all of our Agents are busy, you are currently number two in the queue and your expected wait time is 35 seconds. Please hold until an operator becomes available. Your call is important to us." (The dreaded elevator music then starts playing until the call is transferred to an agent.)
- "Good Day and welcome to XYZ Corporation. So that we can better service your call, please make a choice from the following options. Press 1 for Sales, press 2 for Marketing, press 3 for Technical Advice, press 9 to talk to the Operator, or please hold until your call is answered."
- "Good Day and welcome to XYZ Corporation. Please enter your account number and password followed by the pound sign."

The preceding spoken messages are generally provided to you by an IVR (in this case, the IP-IVR). It is also quite common to have multiples of these prompts combined to provide the caller (whilst they are in the queue) with information relevant to them. Examples of this would be to provide daily cartridge specials to a person who purchases large amounts of printer cartridges. Another option would be, as in the case of the second and third examples, to extract some information from the customer, and use this information to identify their needs, which allows them to pass through to the correct agent the first time. We will be referencing the preceding example messages throughout this IP-IVR section. This

is so we can better assess some of the capabilities of the Cisco IP-IVR and provide an understanding of how the Cisco IP-IVR fits into the IPCC.

The Cisco IP-IVR Solution can run on several different server platforms. Currently, while co-resident with CallManager, the IP-IVR has the ability to be able to service two IVR ports. If more ports are needed, it also has the ability to be able to scale up to 60 ports on a dedicated platform. The actual amount of IP-IVR ports will depend on the number of ports purchased. Currently you have the option to add multiple additional IP-IVR ports as the need arises, but special care needs to be taken to adhere to the maximum number of supported ports as per your IP-IVR hardware platform. Please note that this does not mean that ICM (discussed later in this chapter) will allow you to use all the ports you have purchased. For ICM to control the IP-IVR ports, a separate license needs to be purchased to allow this control to happen.

NOTE

When looking at the different platform options be sure you do NOT underestimate the amount of IVR ports required, but instead cater to the customers' future requirements. The last thing you need is to purchase a platform that caters a maximum of 30 IVR ports and in six months' time realize you need more than the maximum 30 supported ports. As a design rule, try not to exceed 75 percent of the maximum number of ports supported by the platform. Obviously, you may not be able to determine beforehand all the requirements, and in this case, bigger may probably be better (a bigger hardware platform, that is, not an increase in the number of ports).

Before discussing the components of the IP-IVR, it might be best to give you some insight into a few of the capabilities of the IP-IVR. We'll start off by looking at the three different sets of messages, their abilities, and what information is required to make them work efficiently. Let's look at the first example, the one that tells the caller what number they are in the queue.

Looking at this example, there are several variables that need to be considered when using this prompt. For example, how do we know that we are number 2 in the queue? Also, how do we know that the estimated wait time is 35 seconds? The simple answer is that we get the information from ICM (discussed later in the chapter), but if you look at the prompt, you will see the following steps.

- Play the Welcome Prompt File ("Good Day and welcome to XYZ Corporation. We apologize, but currently all of our Agents are busy, you are currently number __")
- Find out the position in the queue from ICM and repeat this to the caller ("2")
- Play an interim prompt in this case ("in the queue and your expected wait time is __")
- Find out the estimated wait time from ICM and tell this to the caller ("35")
- Play the seconds files ("seconds")
- Play another sound file ("Please hold until an operator becomes available. Your call is important to us."). Add music, or possibly an advertisement, or even start the script from the beginning, the choice is yours.

A simple script, like that just mentioned, has several components that need to be recorded separately but joined together to provide a single seamless prompt. Later in this section, we'll discuss how we integrate the different steps of the script.

The second example uses caller input to let them decide which particular department they would like to speak to. This not only saves an agent from having to answer the call, find out the caller's requirements, then do a transfer to the correct agent to deal with the query, but it also helps to increase customer satisfaction as the caller is immediately directed to the correct agents without having to restate any information or be rerouted from the operator to the agent.

Because of the design and flexibility of the IP-IVR, it is possible to reference multiple menus after the caller has made their choice from the initial menu prompts. So, in the previous example, it is possible to play another menu (or multiple menus) to the caller after they have pressed number 1 to go through to Sales. This allows you to further define which department the caller wants to be connected to. Once again, you may give them the option to press 1 to go through to the Router Sales Department, press 2 to go through to Switch Sales Department, press 3 to access the Security Sales Department, press 9 to be connected to an operator, or simply hold for the next available agent.

This last option is normally used in situations where the caller is not able to enter Dual Tone Multi-Frequency (DTMF) tones via the handset, and is normally only needed on the initial menu prompts.

In the third and final example, we take the information the caller has provided (an account number and a personal identification number [PIN]), use it to

make a routing decision. For example, if you have a platinum-level caller, who always spends large amounts of money, and a silver-level caller, who has a small enquiry, calling in at the same time, would it not make sense to provide your big spending customer to take precedence in the queue, and while waiting there, have some relevant information (for example the special of the day) played back to them? This type of flexibility allows you to identify customers based on certain matched criteria and provide a service according to the rules you have set up. These are some of the capabilities available in the IP-IVR. Others include time of day, day of week settings, and so on.

Designing & Planning…

Designing an IVR Script

When a customer wants to integrate an IVR into their contact center solution, normally an IVR is either over- or under-engineered. To help minimize the amount of time spent on IVR configuration, you need to identify the customer requirements, understand how the calls flow through the network, and put a flow chart in place that describes the solution. Once this is done, it will then give you the ability to quickly and efficiently deploy IVR scripts. Also, the IVR is normally the customers' initial access into the network, so it needs to be set up in a manner that is simple and effective, providing customers with an easy-to-navigate, easy-to-listen-to initial contact with your company.

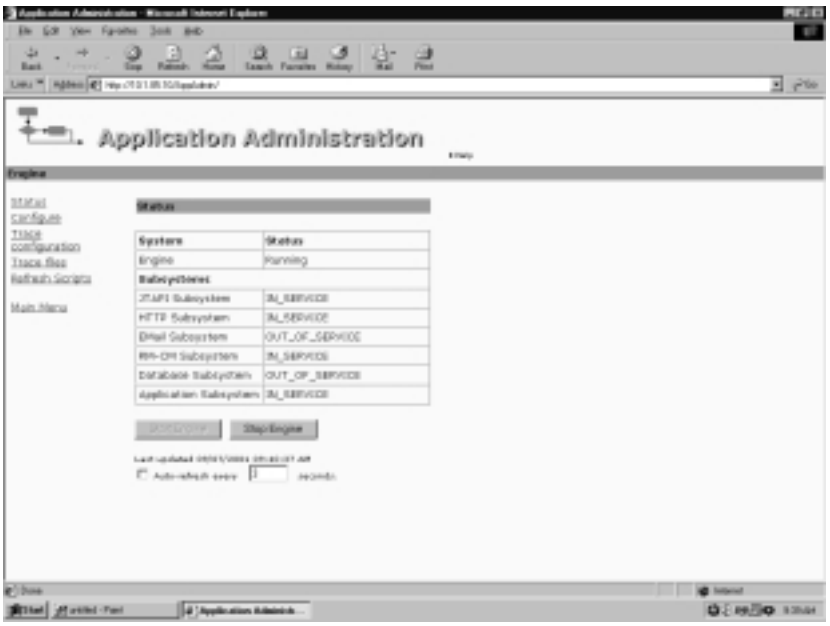
There are two major components to IP-IVR: the Customer Response Application (CRA) Engine, and the CRA Editor. The IP-IVR is just a subset of the CRA's abilities. It is also used in the setup of scripts for the Automated Attendant as well as other types of Cisco-related solutions.

The CRA Editor is a downloadable and installable GUI that allows users to download and edit scripts for any CRA engine regardless of its geographical location in the network. Due to the nature of IP, any of these components can be located anywhere throughout the IP network. The CRA Editor is able to test these scripts (those shown to you in the previous examples), and once tested and customized, upload these scripts to the CRA Engine for use in a live environment. A script is made up of multiple individual steps that should cater to all possibilities

while a caller is in the queue. For example, what happens if the customer does not press any buttons, or presses a button that is either not in use, or is an incorrect choice. All of these eventualities need to be foreseen, and the CRA Editor lets you cater to all these eventualities.

The CRA Engine is the application that runs the scripts you have created or edited with the editor. Within the CRA Engine are multiple subsystems, not all of which are needed depending upon which components are installed. Please refer to the Getting Started Guide to find out more about which subsystems are required, as well as their function. The subsystems that are most relevant are the JTAPI, ICM, and database subsystems. The subsystems control the connections between the ICM as well as the Cisco CallManager, which in an IPCC are critical. Figure 7.1 illustrates the administration of applications for the Cisco CallManager.

Figure 7.1 Application Administration for Cisco CallManager



Cisco Intelligent Contact Management

The Cisco Intelligent Contact Management (ICM) software is probably one of the most important pieces of the Cisco IPCC. Purchased by Cisco, Geotel (as it was called then) offered customers the ability to use a solution that provided integration with multiple vendors: Automatic Call Distributors (ACDs). This let you, via a single interface, pass voice calls through to different vendors' ACDs, which,

in turn, allowed you to have a single set of business rules as well as a single reporting interface that spanned all supported ACDs.

Due to the nature of the solution, it was only a matter of time before Cisco developed an interface into the Cisco CallManager. This interface now provides multiple legacy ACD support as well as gaining from the benefits of integrating with the Cisco CallManager IP-based solution. Via the Cisco CallManager, it allows you to provide customers with a smooth migration path to IP telephony while protecting the existing ACD investments, which are more often than not substantial.

Within the IPCC solution, ICM is the brain that makes the decision on where to pass a customer contact (a voice call, for example) through to. Based on a set of rules held with ICM (explained later in this section), it has the ability to decide which agent or type of agent (service, skill group, and so on) a call should be passed through. If, for example, you have two people enter your contact center simultaneously, you need to be able to provide a predefined level of service to both customers based on certain criteria, like their account number, or the number they dialed. This allows you to categorize customers, for example, according to their current status with the company (Platinum, Gold, Silver). You also need to be able to pass these calls through to the correct agent. In this example, it may be wise to pass the Platinum customer through to an agent before the Silver customer. Also, these rules should make sure a customer is passed to the correct type of agent the first time, negating the need for unnecessary transfers.

NOTE

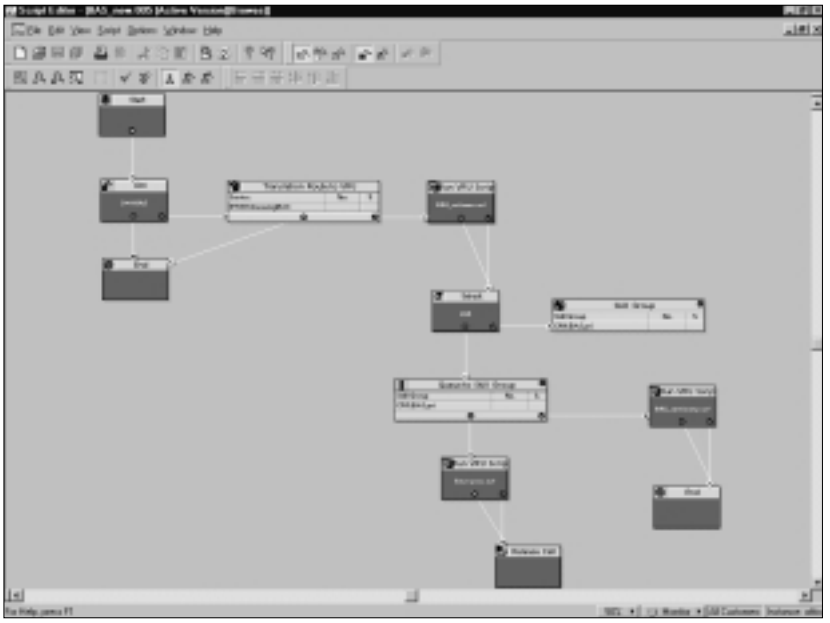
Remember to find out if your legacy ACD is one of the supported integrations. If so, this should ease the move from a legacy solution to one that's IP-based. The list of ACDs as well as PBXs currently supported by ICM is available at www.cisco.com/univercd/cc/td/doc/product/icm/.

With ICM, many different combinations of software are needed, based primarily on customer needs. Some of the more common options will follow shortly. However, these should only be used as indications of some of the components needed. Please consult the Cisco Web site for a listing of partners authorized to see these solutions. One important point to note is that these kinds of solutions should only be deployed with some type of professional services that

make sure the installation is not only done correctly, but supported by Cisco. These partners can be located at www.cisco.com/public/crs/locator/, which provides you with a comprehensive listing of all available partners in your area.

The ICM Script Editor, as the name implies, is where the scripts for ICM lie. ICM uses these scripts (which are usually defined by business processes within an organization) to make decisions on where a call should be sent. Within these scripts it is possible for us to define different groupings of agents based on criteria such as the dialed number or the digits entered on an IVR. These scripts understand that if, for example, the number 555-1212 is dialed, it will take you through to XYZ Corporation's Sales Department, but if 555-9191 is dialed, it will pass the call through to the Service Department. Once the department is defined, we can then decide which set of agents a call can be sent to, and how to send these calls in an even manner. A simple example of this can be seen in the terms Longest Available Agent (LAA) and Minimum Expected Delay (MED). With LAA, a call will be passed to the agent who has been available the longest; whereas MED will pass a call through to an agent that has been calculated to have the minimum expected delay. These types of decisions, as mentioned earlier, should mimic the business processes already in place, as well as be an extension of them. Figure 7.2 shows the ICM Script Editor Interface.

Figure 7.2 ICM Script Editor Interface



IPCC Hardware and Underlying Infrastructure Requirements

The underlying hardware and infrastructure requirements to provide optimal IPCC performance are often overlooked. While we won't go into too much detail here, certain things need to be taken into consideration.

Hardware components for CallManager as well as IP-IVR are fairly straightforward since the only choices are on Cisco-provided or customer-provided platforms. Should you choose the customer-provided platform, however, check the Cisco Web site to make sure the specific product is supported.

ICM is slightly more difficult since multiple items may be required depending on the customer's needs. For a comprehensive listing, refer to the following Cisco Web page: www.cisco.com/warp/public/78/hardwr_specs.html. All the optional components (E-mail Manager, Collaboration Server, and so on) as well as their hardware requirements are also listed on that page.

Cisco routers, or gateways, are discussed in Chapter 3 (AVVID Gateway Selection). Quality of Service, another overlooked component, will be covered in Chapter 8. Based on this, the underlying infrastructure (switches and so forth) can be chosen.

Providing Voice Recording Options

Cisco currently does not provide a VoIP voice recording solution, but rather relies on technology partners to provide these types of integrated solutions. This section will discuss the various VoIP solutions available.

One important point to remember is that VoIP voice recording differs from traditional voice recording solutions. In a VoIP solution, we talk about bits and bytes, as well as things like Real Time Protocol (RTP) streams, and H.323 and Skinny Protocols. These types of terms require a completely different mindset. Previously, it was fairly straightforward to add a voice recorder, since it basically taped the conversation coming into the PBX. Unfortunately, we now have no central point where we can place a voice recording solution, and even if we could, it would be garbled because traditional solutions cannot understand and decode IP streams (unless, of course, they were placed *before* the conversion to IP).

An easy way to think about this is to compare the following voice recording scenarios. First, let's look at a normal cassette recorder. Recording from a radio station onto a cassette is easy. It's a simple matter of placing a tape into a cassette recorder, pressing record, and voilà, you're recording from radio to tape.

Now, take the same solution, but apply it to an Internet radio station. Questions arise. Where do you place the cassette recorder, and how exactly do

you record? You might respond, "I'll put it next to my speakers and record there," but you'd still be relying on the PC's ability to decode the IP packets coming in from the Internet. Plus, sound quality would be poor.

The same goes for IP recording. While you can take a traditional solution and crudely adapt it, vendors providing IP-based recording offer better solutions. But how exactly is IP voice recording accomplished, and what should be taken into account when looking to implement a voice recording solution?

As mentioned previously, it is possible to apply a traditional solution to voice recording. This is achieved by placing the voice recording solution *before* any of the conversations have been packetized, an option that allows *all* voice conversations on a T-1 link to be recorded. But a problem arises in that there are no mappings as to who the call actually went to. We may have a time the call recording started, but given that we have hundreds of agents, we won't know to whom the call was transferred. Tracking down a particular call would be very laborious and would probably require a client to listen to many recorded voice calls. As mentioned previously, a good system will log *all* voice conversations, which essentially means you will need a voice recording method with a large amount of storage space.

This is why a solution has been created specifically for IP telephony solutions, and why specific requirements need to be met before implementing these solutions. First of all, let's discuss how to record these sessions, and how to decode and provide Administrative information about the voice calls.

Current solutions rely on a feature available on the majority of catalyst Ethernet switches, called Switched Port Analyzer (SPAN). SPAN mirrors traffic coming in on one or more defined Ethernet switch ports and passes this information to the SPAN port. This port is normally used in areas of network management or when doing some type of packet decoding (which is done when we utilize the VoIP voice recording solutions). The following is an example of how the switch is configured:

```
Cat> (enable) set span 2/1 2/2

Enabled monitoring of Port 2/1 transmit/receive traffic by Port 2/2
Cat> (enable) show span

Destination      : Port 2/2
Admin Source     : Port 2/1
Oper Source      : Port 2/1
Direction        : transmit/receive
```

```
Incoming Packets: disabled
Console> (enable)
```

To simplify matters, we need to put a VoIP recording application somewhere where it can "hear" the conversations going on in the network around it. Because the application receives all the IP packets, it can then make a decision (based on configuration) as to whether the packets should be decoded, and where to store them on the voice recording application server along with the necessary administration information. While this is probably good enough for the majority of the IPCCs, there are some limitations you should be aware of.

Firstly, you are limited by the speed of the SPAN Port. If it is only a 10 Mbps port (normally it's 100 Mbps), you may have problems sending all the information to the single SPAN port.

Secondly, you must be able to configure a SPAN port (or something similar) on your Ethernet switch. Without this, there is no non-intrusive way in which you can monitor the voice traffic traveling your network.

Lastly, during a call, whenever ten seconds of silence occurs, the call normally ends. However, if you were using an older version of CallManager (which did not provide Music on Hold), and an agent put a call on hold while they consulted another agent, , for every ten seconds the caller was on hold, a new call would be generated. So, if the caller were put on hold for two minutes exactly, we would actually end up with 14 voice recording calls (1 to start, plus 12 x 10 second intervals calls, and another to finish the conversation). Obviously, we would not want this. To get around this problem, CallManager 3.1 and later do provide Music on Hold, allowing you to have a single conversation voice recording.

A workaround may be to utilize some of the capabilities of the catalyst switches, which give you the ability to be able to configure your IP phone in one virtual LAN (VLAN), and your PC in another VLAN. This would then allow you to be able to SPAN a VLAN (as shown next), which in turn means the information you are receiving via your SPAN port is *only* voice-related.

```
Console> (enable) set span 6 2/1

Enabled monitoring of VLAN 6 transmit/receive traffic by Port 2/1
Console> (enable) show span

Destination      : Port 2/1
```

```
Admin Source     : VLAN 6
Oper Source      : Port 1/1-2
Direction        : transmit/receive
Incoming Packets: disabled
Console> (enable)
```

There will be mechanisms whereby these types of voice recording solutions will not have to use the SPAN port of an Ethernet switch, but will be agent-initiated. These solutions will probably cater to H.323 capabilities, as well as silent conference setups. Although we term it a conference call, in essence, there is only the agent, the caller, and the third party, which is the voice recording solution. When using the agent-initiated option, this is probably how the call would be set up. Currently, two vendors are providing solutions that integrate with the Cisco IPCC:

- **Nice** www.nice.com
- **Eyretel** www.eyretel.com

For more information on up-to-date voice recording solutions as well as other IPCC partners, visit www.cisco.com/warp/public/180/prod_plat/cust_cont/ipcc/part_doc.html.

Call Accounting, Billing, and Network Management Solutions

As with all companies, there is a bottom line. Many times, a justification of expenditures is required, so the company can decide whether a particular solution is helping both them and their customers. To analyze how their own company is benefiting, they may look for a return on investment (ROI) and try to track the income from the installed solutions. Your customers (if this is your own company) will look at the billing architecture and structure to see if they are receiving a good return on their investment as well. To measure these metrics at the company level, however, call accounting, billing, and network management options must be implemented.

Call Accounting and Billing Solutions

TMSs or Telephone Management Systems are an integral part of any telephony network, and irregardless of whether you're using a traditional telephony solution

or an IP telephony solution, the guidelines stay the same. When implementing a call accounting or billing solution within an IP telephony environment, the same values you would look for in a TMS when used in a traditional solution will be applicable in an IP telephony solution. Next, we'll discuss some of the factors to take into account when looking to implement a call accounting and/or billing solution.

In an IP solution, the idea of call accounting and billing is based on a Call Detail Record (CDR), which, as the acronym suggests, is a detailed record of a call. Currently, the Cisco CallManager has the ability to provide CDRs, which can be created on certain Cisco VoIP gateways. Used in conjunction with the solution mentioned earlier, we can achieve both call accounting on a local and global scale. Before getting into any guidelines regarding the implementation of these applications, it might be best to give you an idea of the information contained within both the Cisco CallManager CDRs and the VoIP gateway CDRs.

The Cisco CallManager, as mentioned previously, is a Windows 2000 Server, that comes with SQL Server as well as the relevant service packs installed. Using SQL, the Cisco CallManager has two databases that store information with regards to a call:

- **CallDetailRecord** This database collects call-specific information.
- **CallDetailRecord Diagnostic** This database contains values which are used in conjunction with applications to determine the quality of the call commonly known as a Call Management Record (CMR).

NOTE

There may be multiple entries in the CallDetailRecord Diagnostic Database that relate to a single entry within the CallDetailRecord Database due to certain factors such as quality changes, and so forth.

For a call accounting/billing application to pull data from a Cisco CallManager or CallManager cluster, CDRs first need to be enabled. This is achieved by launching the CallManager Administration tool from within a Web browser and changing the values of the following two parameters, both of which are disabled by default. If you have multiple CallManagers within a cluster, make sure both parameters are changed on all the CallManagers, otherwise no CDRs will be available (see Figure 7.3).

Figure 7.3 Cisco CallManager Service Configuration

Once this process is completed, every time a call is placed (even for uncompleted calls), a CDR is generated. If there is only a single CallManager, the CDR will obviously be created on this single Server; however, if there are multiple CallManagers making up a cluster, there will be both a publisher and subscribers. Based on which CallManager within the cluster is the publisher (this is defined when installing the CallManager software), the CDR will be created on the publisher alone.

Cisco provides a solution as an add-on that can be downloaded from www.cisco.com. This solution is called Cisco Administrative Reporting Tool (ART), and can be used to provide basic billing as well as the QoS on the network. Looking at the screenshot that follows, it is possible, via this tool (which is accessed from the CallManager main screen), to provide reports in PDF format that can give you, for example, the billing per individual for a specific date range. ART does have its limitations, however. For instance, it can only handle up to 2,000,000 CDRs. This may seem like a lot, but remember, for each inbound or outbound call, information must be logged. To turn this idea into an example:

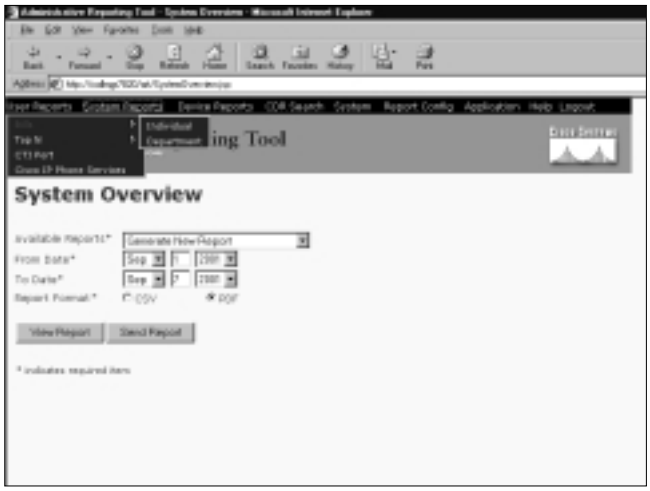
Suppose there are 250 employees in Company XYZ, who make and receive, on average, around 15 calls per person per day. To define how long ART would last, you would have to perform a calculation as follows:

250 (employees) x 15 (calls per day) = 3750

2000000 (max CDRs supported by ART) / 3750 (CDRs per day) = 530

In the preceding example, it is clear we have enough capabilities within ART to support up to approximately 530 days worth of CDRs. Again, this may sound like a lot, but if you run a call center, or have dialing campaigns, you will need to work out the maximum number of CDRs per day (see Figure 7.4).

Figure 7.4 Cisco CallManager Administrative Reporting Tool

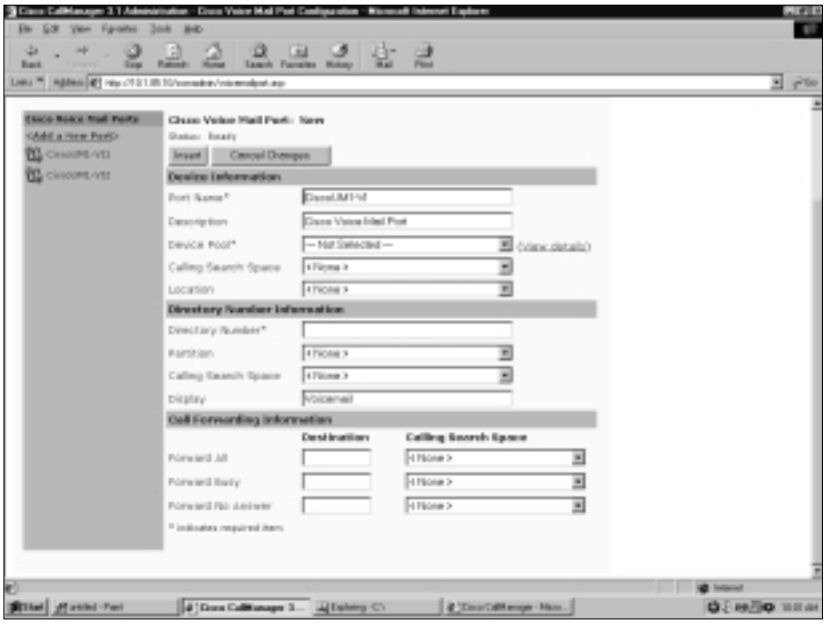


Designing Voice and Unified Messaging Solutions

Cisco has been providing voice mail solutions for the IP telephony market for a while now. In earlier releases of Cisco CallManager, a voice mail solution was included as a free co-resident option on the Media Convergence Server (MCS) Server Platforms. This product, uOne, has now been discontinued, and a migration plan has been put in place by Cisco to move over to their new voice mail/unified messaging solution which is called Unity. Based on this, this section will focus solely on Cisco Unity, making the assumption that either the migration to Unity from uOne has taken place or you are using other solutions. (Examples of solutions that integrate with Cisco CallManager can be found on the Cisco Web site.)

In keeping with Cisco's strive to make their IP telephony and application solutions as open and standards-based as possible, members of the Cisco AVVID Partner Program provide Unified Messaging Solutions that are in the process of being integrated into the Cisco IP telephony solution. This essentially means that it is possible to purchase the IP telephony solution from Cisco, yet buy (or integrate an existing) unified messaging solution into your CallManager network. Some of the vendor's solutions will be discussed later in this section. As it happens, one of these integrated solutions was a product called Active Voice Unity, which Cisco purchased in late 2000, and which was subsequently integrated into the Cisco fold in early 2001. Cisco Unity 2.46 was the first version released by Cisco, currently we have in the 3.*x* range. One of the advantages of using the Unified Messaging Solution with Cisco CallManager is the simplicity of the CallManager configuration itself (see Figure 7.5).

Figure 7.5 Adding Cisco Voice Mail Ports



Before going further, it might be best to give you an overview of what Cisco Unity is, and provide some simple guidelines as to its use. Cisco Unity is currently available in two different options. The first is voice mail only, and the second unified messaging, which includes the voice mail option (we'll discuss the differences further on). Keep in mind, if you end up using voice mail only, you

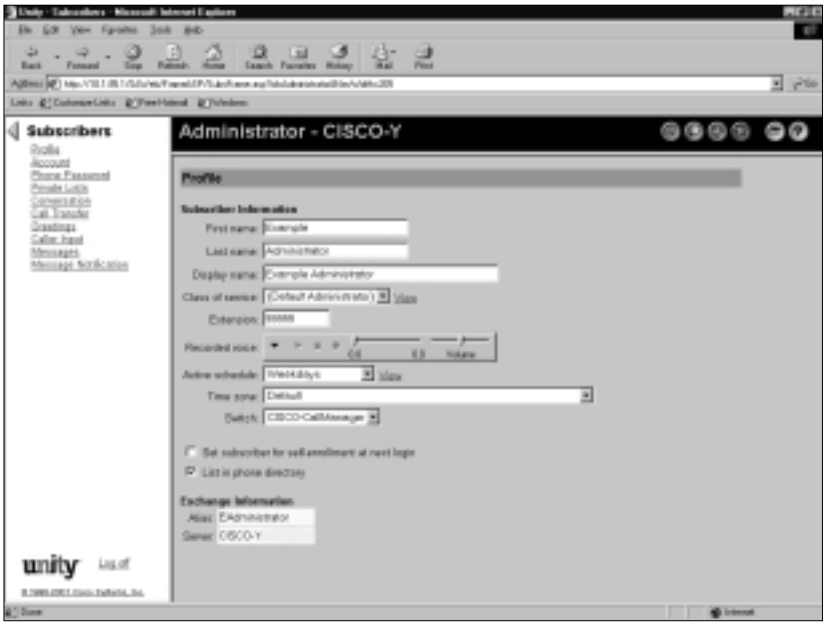
have the choice of upgrading later to the unified messaging option (see Table 7.1 and Figure 7.6).

Table 7.1 Services in Cisco Unity Solutions

Options	Cisco Unity Voice Mail	Cisco Unity Unified Messaging
Voice Mail	X	X
E-mail		X
Fax		X
Text To Speech		X
Multiple Directories *	X	

**Cisco Unity voice mail can have multiple directories and message stores. There is a minimum of two, one for voice mail, and another for e-mail. If there is no fax integration, there may even be a third store. Cisco Unity unified messaging, meanwhile, supports a single directory as well as a single message store.*

Figure 7.6 The Administration of a Cisco CallManager



Other points to bear in mind when looking at the Unity solution is the scalability as well as the integration into legacy PBXs. Cisco supports certain legacy PBXs (Nortel Meridian 1, for example) allowing you a smooth transition to IP telephony. A listing of those currently supported can once again be found on the Cisco Web site.

WARNING

If you're looking to migrate from voice mail only to unified messaging, remember to make sure the hardware requirements for running unified messaging is met. Even though it is only a software upgrade, the requirements are not the same.

Understanding Other Voice Applications

To provide you with an understanding of other applications available when using the Cisco AVVID architecture, let's consider specific solutions you're likely to see in your daily work. These are not the only solutions available, but they are some of the most common.

IP Automated Attendant (AA) is the first application we'll discuss. It is extremely useful in large organizations where switchboard operators are normally overworked. Automated Attendant, as its name suggests, provides automated functions an attendant might normally perform. Mundane tasks, such as transferring a call to an extension, do not need to be done by an operator. Now, by prompting a caller as to whether they know the extension of the person they're trying to reach, or even the person's surname, the call can be forwarded to that extension without the intervention of an operator. Within these scripts (the same type used by the IP-IVR), the caller can, of course, always be connected to the operator if they don't know the information they're being asked.

AA can be run co-resident with Cisco CallManager, but it has limited scalability, only allowing a maximum of ten ports. If more ports are needed, it can be run on a separate server, but as with all of these applications, there are guidelines available on the Cisco Web site that state the current maximum number of sessions relevant to the hardware platform purchased.

Although AA may sound like an IP-IVR, there are some differences. The IP-IVR is more customizable, and has more features. If call answering and distribution

is what is needed, then AA is more of a solution. However, if a queue point as well as other IVR-type features are required, the IP-IVR should be used.

Similarly, the Cisco WebAttendant, like AA, is meant to ease an operator's workload. WebAttendant is a Web-based graphical user interface (GUI) that works with a standard Web browser without making any changes to the browser itself. The only thing needed for the installation is to download the application from the Cisco CallManager Install Plug-ins page.

Cisco Conference Connection, meanwhile, is a software solution that provides additional conferencing capabilities. Installed on a Media Convergence Server, it scales far beyond the capabilities of Cisco Call Manager. It can also be seamless integrated with Legacy PBX and PSTN. The Cisco Conference Connection is a meet-me conferencing solution, which can essentially schedule a conference where multiple parties dial a specific number at a specific time and, using a Cisco IP phone, conference other members in.

Cisco IP phone productivity services allows any user utilizing a Cisco 7940 and Cisco IP Phone 7960 to view/listen to their voice mail/e-mail as well as access their online personal calendar. This essentially means that a person with the correct type of IP phone does not need to have a PC powered up to view e-mail or see what their calendar looks like. In conjunction with this solution, there is an IP phone services Software Development Kit (SDK) which allows developers to create their own applications to the supported IP phones.

Summary

Applications are the differentiating factor for IP telephony solutions. If you have a customer that only picks up the phone, dials a number, speaks, and then puts the phone down, it is difficult to explain the differences between IP telephony and traditional solutions, basically because both (at least in that scenario) do exactly the same thing. The applications that IP telephony offers currently, as well as some of the advantages possible in the near future, not only make applications more advantageous than traditional solutions, but more open. We discussed how open-based standards support has been created, and how Cisco adheres to this through interfaces like TAPI and JTAPI, which are supported via IP telephony solutions that allow other parties to integrate proven solutions into them.

We talked about IVR and how it can provide initial customer interaction by providing the caller with relevant information while waiting in the queue. This allows the caller to do self-service type queries (check their bank balance and so on), as well as access menu-driven options which allow them to define the type of person or department they wish to speak to. This is also in keeping with the idea that IP telephony should be location-independent allowing administrators to place solutions where they like, as opposed to being in a computer center. Though platform-dependant, the system can scale up to 60 ports.

We also discussed the use of WebAttendant, a Web-based GUI, downloadable from the CallManager Install Plug-ins screen. This software-based attendant solution negates the need for operator consoles to have dedicated and specific handsets they can use. Also, once again, due to the nature of IP telephony, not only can we have many WebAttendants throughout the network, but they don't even need to be in the same country when answering calls.

Implementing the IP contact center solution, which is made up of several components such as CallManager, IP-IVR, and ICM allows you to keep your investment in existing technology while facilitating the quick deployment and movement to IP telephony. Certain guidelines are mentioned at the beginning of the IPCC, and these guidelines MUST be adhered to in order to succeed in deploying the IPCC solution.

We also discussed unified messaging and how this type of solution is just emerging, with its ability to manage all points of contact from a central space. It allows us to listen to e-mails and view voice mails without ever having to power up the PC. Its parent program, Unity, also permits us to scale up to 5,000 mailboxes on a single server.

We talked about the Personal Assistant and how it can be used to define the rules by which you receive communications. Essentially, it allows you to use speech recognition from any handset as well as determine whether you want to receive calls based on a set of individual rules, and then pass these calls through to a specified route. This then lets you be more efficient in the way you conduct business, since you receive only the calls you want, when you want them.

Phone productivity services, meanwhile, provide you with a single access point, via the Cisco large display phone, to look into your corporate e-mail, voice mail and diary. If the need arises, you also have the ability to sync with your personal contacts, once again providing a single point of contact for communication.

Finally, there is the Cisco AVVID Partner Program. This is probably the most important point, as it gives you the ability to choose the partner with whom you would like to integrate. Because of the nature of the CallManager and other components, it allows you to provide a simple interface into other vendors' solutions. Please consult the Cisco Web site for the latest listing of supported vendors' solutions.

Solutions Fast Track

Creating Customer Contact Solutions

- ☑ Make sure you understand the customer's needs.
- ☑ Provide the client with the solution that best suits these needs.
- ☑ Make sure to stay within the Cisco recommended guidelines.
- ☑ With the IP contact center, there are many different components. Make sure the version numbers needed to run the solution are all compatible.

Providing Voice Recording Options

- ☑ Make sure the infrastructure can support voice recording.
- ☑ Define the endpoints that need to be recorded, and implement a policy using this as a framework.

Call Accounting, Billing, and Network Management Solutions

- ☑ Understand the requirements in enabling CDRs throughout your network, not just on the Cisco CallManager, but also on your router infrastructure (if possible).
- ☑ Look at the Administrative Reporting Tool (ART) with Cisco CallManager to decide whether this would provide you with the information needed before looking at external solutions.
- ☑ Define the information needed with your reports, and based on this, look for solutions that meet the requirement you and your customers have.

Designing Voice and Unified Messaging Solutions

- ☑ Decide on the version of Unity needed.
- ☑ If upgrading from voice mail to unified messaging, do not forget the possible hardware requirements.
- ☑ You should be running Microsoft Exchange 5.5 or Exchange 2000, with future support for other platforms.

Understanding Other Voice Applications

- ☑ Keep it as simple as possible, if services or applications are not needed, do not enable them. It complicates the configuration.
- ☑ IP Automated Attendant (AA) is extremely useful in large organizations where switchboard operators are normally overworked. Automated Attendant, as its name suggests, provides automated functions an attendant might normally perform.
- ☑ WebAttendant is a Web-based graphical user interface (GUI) that works with a standard Web browser without making any changes to the browser itself. The only thing needed for the installation is to download the application from the Cisco CallManager Install Plug-ins page.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: What is the maximum amount of IP phones per single CallManager, and the maximum amount per CallManager cluster?

A: The maximum amount per CallManager server is currently 2,500 IP phones, with the maximum per CallManager cluster currently limited to 10,000 IP phones.

Q: What components are essential for an IPCC?

A: The underlying infrastructure needs to be able to support voice solutions, along with Cisco CallManager, Cisco IP phones, Cisco IP-IVR, Cisco ICM, Cisco CTI Solutions/Agent Desktop, and server hardware requirements.

Q: Is text-to-speech available on the Cisco Unity voice mail component?

A: No, this capability as well as others mentioned earlier in the Unified Messaging section is only available on the Cisco Unity unified messaging component.

Q: When should ART be used?

A: When you will not be exceeding the maximum amount of CDRs supported, and don't need the comprehensive reports that a dedicated TMS would provide. Another reason for its use would be to measure basic voice latency on your network.

Chapter 8

Advanced QoS for AVVID Environments

Solutions in this chapter:

- Using the Resource Reservation Protocol
- Using Class-Based Weighted Fair Queuing
- Using Low Latency Queuing
- Using Weighted Random Early Detection
- Using Generic Traffic Shaping and Frame Relay Traffic Shaping
- Running in Distributed Mode
- Using Link Fragmentation and Interleaving
- Understanding RTP Header Compression

- ☑ Summary
- ☑ Solutions Fast Track
- ☑ Frequently Asked Questions

Introduction

This chapter outlines some of the newer and, therefore, more AVVID-related QoS mechanisms currently available in Cisco IOS. While some of these mechanisms are just beginning to be widely deployed on production networks, QoS is constantly being improved, expanded, and modified, so these technologies will undoubtedly continue to receive a significant amount of development effort in the near future. Of course, entirely new QoS mechanisms will continue to emerge as Cisco's AVVID strategy pushes forward into a growing number of customer networks as demand increases for even more advanced QoS controls. Table 8.1 provides an overall comparison between some of the different QoS mechanisms we will discuss in this chapter, including: Resource Reservation Protocol (RSVP), Class-Based Weighted Fair Queuing (CBWFQ), Link Fragmentation and Interleaving (LFI), compressed Real-Time Transport Protocol (cRTP), and Low Latency Queuing (LLQ).

Table 8.1 Comparison of QoS Mechanisms

QoS Mechanism	Scales to Large Networks	Supports High-Speed Interfaces	Designed to Support Voice
RSVP	No	Yes	Yes
CBWFQ	Yes	Yes	No
LFI	Yes	Use on Links < 768k	Yes
cRTP	Yes	Designed for Slow Links	Yes
LLQ	Yes	Yes	Yes

Several of these technologies, such as RSVP and LLQ, are currently being used for voice applications (mostly), and you will find that these more advanced mechanisms are often used in conjunction with each other, rather than independently. Such mechanisms, although powerful and useful in their own right, gain power and functionality when used alongside other mechanisms. LLQ, for example, is simply Class-Based Weighted Fair Queuing with the addition of a Priority Queue. For this reason, implementing LLQ in your AVVID environment will necessitate the use of CBWFQ even if you simply dump all your nonvoice traffic into the single default class (class-default).

Some of the benefits of more advanced queuing mechanisms include increased granular control of traffic behavior, and the ability to be far more specific when classifying and queuing, or dropping, traffic. However, this presents a

potential problem. There is a trade-off between granular control and flexibility of use. LLQ, for example, is a very specific mechanism with a very specific purpose, but it is not well suited for many things other than that particular function. This is a contributing factor to the reasoning behind using multiple QoS mechanisms to develop a strong overall Quality of Service strategy for your AVVID environment. It is particularly important in this chapter to pay attention to recommendations about where the deployment of these mechanisms is appropriate.

Using the Resource Reservation Protocol

The Resource Reservation Protocol (RSVP) is the first attempt at an industry standard implementation of the Internet Integrated Services (Intserv) model of QoS. Researchers at the Information Sciences Institute (ISI) at the University of Southern California (USC) and the Xerox Palo Alto Research Center first conceived of RSVP.

NOTE

In 1993, the Internet Engineering Task Force (IETF) started working toward standardization through an RSVP working group. Version 1 of this protocol is currently defined by RFC 2205. Interested readers may find the IETF Applicability Statement (RFC 2208) helpful in pointing out both the uses and current issues with an RSVP deployment. This chapter will illustrate both of these briefly.

The Intserv model is characterized by applications or end stations reserving resources across a network to guarantee a particular level of service. RSVP is a protocol that implements this signaling.

RSVP is independent of, yet complimentary to, Intserv. Whereas Intserv specifies the set of classes of service and the parameters to deliver QoS, RSVP requests this service level from network nodes and, in some cases, carries the Intserv parameters.

RSVP does not provide QoS directly to applications, but instead, coordinates an overall service level by making reservation requests with as many nodes as possible across the network. It is up to other QoS mechanisms to actually prevent and control congestion, provide efficient use of links, and classify and police traffic. A successful implementation of RSVP requires that it work in conjunction with these other mechanisms.

RSVP's popularity lies in its capacity to give guaranteed QoS to real-time applications, such as voice, that have either constant bandwidth requirements or low latency requirements. This is why its primary use today on networks is to deliver multimedia streams such as voice and video, making it a viable option for your overall AVVID QoS strategy. Unfortunately, there are some scalability concerns with RSVP, which we will address as the chapter progresses.

What Is RSVP?

RSVP is a signaling protocol that makes reservations of resources (bandwidth and so on) for client applications to guarantee a certain QoS. It is considered a signaling protocol because these reservations are negotiated by communication between end stations. Furthermore, it is an out-of-band signaling protocol. RSVP packets are not used to transmit bulk data; they coexist on the network with other packets and are used to reserve resources for typical IP packets or, more specifically, the IP packets that make up the flows that are to get specialized QoS. For this reason, RSVP seems like a natural choice when implementing AVVID solutions—since your voice and video traffic have specific requirements, including that for bandwidth—whereas your data traffic might be a bit more forgiving.

RSVP makes reservations of resources for data flows across a network. These reserved flows are usually referred to as *sessions*. A session is defined as packets having the same destination address (unicast or multicast), IP protocol ID, and destination port. Resources could be considered such things as allocated bandwidth, CPU cycles, or queuing priority. Clients use RSVP to request a guarantee of QoS across the network. Routers participate in RSVP by allocating resources to particular flows, or denying resources if there are none available, and by forwarding RSVP information to other routers.

Designing & Planning…

Why RSVP Has Scalability Issues

A signaling protocol can be either in-band or out-of-band, depending on whether the signaling data flows over the same medium as the bulk data. In the telephony world, Integrated Services Digital Network (ISDN) would be considered an out-of-band signaling protocol, because all information for setting up a call passes over a separate D-channel (*data*),

Continued

whereas the actual telephone conversation flows over the B-channel (*bearer*). In a way, RSVP could be considered an in-band signaling protocol, since it flows over the same physical media, the same data-link layer, and even the same network as the bulk data. However, it is usually referred to as out-of-band because the packets are separate from the bulk data. A flow that was set up by RSVP would have nothing in its packets to indicate that it is participating in RSVP. The state of active reservations is stored in each routed node.

As we indicated earlier, RSVP has some scalability issues. To accomplish truly guaranteed resource reservation end-to-end, the state of active reservations must be stored in every router along the path of the packets, in which case a single reservation takes up resources on every router in the end-to-end path. Even worse, the reservations are unidirectional, so a reservation must also be stored in every router end-to-end for the return path of the packets in a bi-directional transmission (such as a phone call).

NOTE

Improved scalability may be realized by not enabling RSVP on every router end-to-end. While we mentioned that true, end-to-end resource reservation can only be accomplished by using RSVP on every router end-to-end, it is not mandatory that RSVP be enabled everywhere on a network. RSVP has the built-in capability to tunnel over non-RSVP aware nodes (see the discussion later on the setup process). Though a truly guaranteed QoS will not be possible in this case, provided that the non-RSVP network has sufficient bandwidth (for example, tunneling over a Gigabit Ethernet or ATM core), the solution might be feasible for most applications. In the case of voice in an AVVID environment, this solution should be tested thoroughly before considering it for production environment deployment. In addition, it is not necessary for the clients to be RSVP-capable. Cisco routers provide RSVP functionality for Voice over IP (VoIP) dial peers. This is accomplished by using RSVP proxy—a function that emulates clients sending RSVP Path and Resv messages.

RSVP is not a routing protocol, it is an Internet Control Protocol that resides at Layer 4 of the Open Systems Interconnect (OSI) reference model, the transport layer. It is similar to other control protocols, such as Internet Control

Message Protocol (ICMP) and Internet Group Management Protocol (IGMP). It is fully compliant with Internet Protocol version 4 (IPv4) and the emerging Internet Protocol version 6 (IPv6). The path that it takes across the network is the same as other IP packets and is determined by the underlying routing protocol (Open Shortest Path First [OSPF], Enhanced Interior Gateway Routing Protocol [EIGRP], Border Gateway Protocol [BGP], and so on).

Besides interoperating with routing protocols, RSVP also works with QoS implementation mechanisms. These are the mechanisms that provide QoS directly, such as weighted fair queuing (WFQ), weighted random early detection (WRED), and the like. What implementation mechanisms are used is not the direct concern of RSVP. It is up to the routers to determine how QoS will be implemented, based on their own particular capabilities. RSVP only makes the request and leaves it up to the intermediary nodes to deliver QoS.

Both unicast and multicast traffic are supported by RSVP. Support for multicast is fortuitous, since RSVP is currently used the most prevalently for voice and video traffic, much of which is characterized by multicast flows. We will see later how RSVP interoperates with the IGMP and Protocol Independent Multicast (PIM) to reserve resources for multicast flows.

What RSVP Is *Not*

As mentioned before, RSVP is not a routing protocol. It relies on typical IP routing protocols to forward the RSVP packets. The next section shows how RSVP uses the routed path to create the setup messages that make the actual reservations.

Because of its protocol-based nature, RSVP does not monitor reservations. It is, therefore, not a resource manager. It is worth reiterating that it is simply a signaling protocol—client talking to client, router talking to router. It does not actually control what kinds of resources are reserved, either. That is up to the routers and their particular capabilities. You can imagine the benefit to a network administrator of knowing at any given moment how many reservations are made across the network. This would help for bandwidth planning and provisioning. Although it is possible to see what reservations are active in the routers, as we will see in Chapter 9, the Resource Reservation Protocol has no capability of providing this information directly.

Although RSVP is an important QoS mechanism, it is not an implementation mechanism. It could be better thought of as a mechanism that requests QoS from other mechanisms. It is not a packet scheduler, link-efficiency mechanism,

or traffic classifier. It does, however, work with these mechanisms. Otherwise, there would be no actual QoS—just a reservation!

How Does RSVP Work?

Now that we have a basic understanding of what RSVP is and is not, let us look at the mechanics of setting up a RSVP session, say, in the case of a phone call or video conference. We will not focus specifically on configuring RSVP but, rather, we will concentrate on the overall strategy.

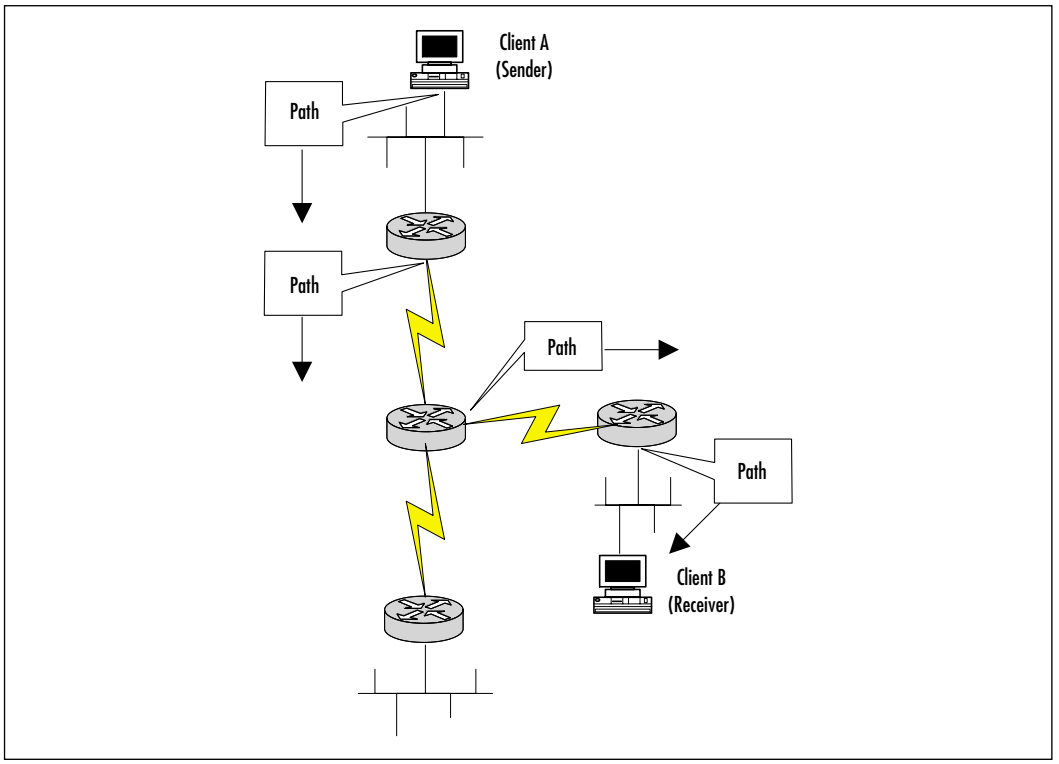
Session Startup

RSVP is often set up between two clients in a point-to-point situation (such as a phone call), or between multiple senders and multiple recipients (multicast). It is even possible for RSVP to negotiate a multipoint to single-point transmission. In any case, the RSVP session startup process reserves resources in a single direction only. To have full-duplex QoS guarantees, it is necessary for the session startup process to be performed twice, once in each direction. For example, in the case of setting up a VoIP call between two telephone users, it would usually be necessary to set up two reservations, one each way, to guarantee good QoS for both callers. On the other hand, a video stream would necessitate only a one-way reservation. Since we are talking about RSVP in an AVVID environment, however, let us consider the reservations required for a videoconference between two people. Since we know that the voice and video components have different bandwidth requirements, there would obviously need to be a separation of the reservations for voice and video. Consider that both pieces (voice and video) would need to be bi-directional, and now we have the need for a total of four reservations between any two given routers. If you take that example and apply it to a four-point any-to-any videoconference, you have 4 x (4 - 1) = 12 reservations being managed on each router. When using the formula A x (B - 1) = C, where A = Bi-directional Flows, B = Total number of Routers, and C = Total Reservations per router, it isn't difficult to realize the challenges you will face when trying to scale RSVP in a large-scale production environment. This is especially true in hub and spoke environments where all videoconferences between remote sites must pass through a hub router (which, in order to guarantee true end-to-end QoS, must maintain reservations for each flow).

Now that we've explored some of the background information that you'll need to keep in mind when considering how many sessions will be started, let us step through the process of an RSVP session startup. In Figure 8.1, we have two

AVVID clients across a RSVP-enabled network (perhaps, a desktop-to-desktop videoconference). At the top we have Client A, which we will designate as the sender, and at the bottom we have Client B, which we will consider the receiver. Thus, after the reservation is set up, the data, whether it is voice, video, or something else, will flow from Client A to Client B in a downstream manner.

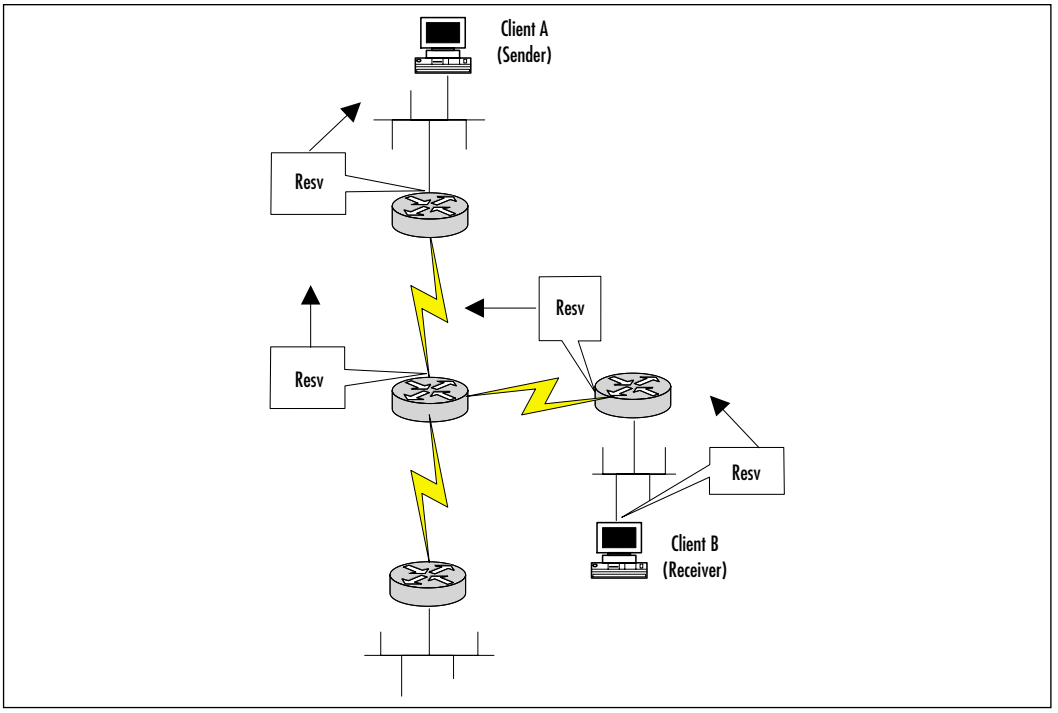
Figure 8.1 RSVP Session Startup, Path Messages



The first step is for Client A, the sender, to transmit a RSVP Path message to Client B, the receiver. This Path message travels across the network according to the underlying routing protocol. At each hop through the network, the Path message is modified to include the current hop. In this way, a history of the route taken across the network is built and passed to the receiver, Client B.

Now that Client B has the complete route from the Path message, a reservation (Resv) message is constructed and sent to Client A along the exact reverse path, as shown in Figure 8.2. At each hop, the router determines if the reservation can be made, based on available bandwidth, CPU cycles, and so on. If the reservation is possible, resources in the router are allocated, and the Resv packet is forwarded upstream to the previous hop, based on the information in the Resv packet.

Figure 8.2 RSVP Session Startup, Resv Messages



NOTE

In both the Path and Resv messages, the upstream hop is usually referred to as the *previous hop,* and the downstream hop is called the *next hop*. This terminology is derived from the reference point of the data moving in a downstream direction, from sender to receiver.

If the reservation is declined, an error message is sent to the receiver, Client B, and the Resv packet is *not* forwarded. Only when Client A receives a Resv packet does it know that it can start sending data and guarantee a particular QoS to the downstream receiver, Client B.

You may think it is odd for the entire RSVP process to begin with the sender building the Path message to the receiver. This might be analogous to a television network deciding it is time for you to watch your favorite show and automatically turning on the TV. However, there is usually some kind of non-RSVP request originating from the receiver to set up this flow. Some good

examples of this type of request might be an H.323 conversation between IP telephony applications, or an IGMP request to join a multicast group to watch a video clip.

NOTE

Though it is necessary for the sender to first transmit the Path message before the receiver can transmit the Resv message, RSVP is still considered receiver-oriented. That is, the receiver of the data flow initiates and maintains the actual resource reservation used for that flow.

Session Maintenance and Tear-Down

After a session is initiated, it is maintained on the routers as a *soft state*. With a soft state session, the path connecting two end stations can be renegotiated without consultation with those end stations. This contrasts with a circuit-switched network, where the connection between two end stations is a hard connection, and when a failure occurs, the connection is broken.

This soft state session must be refreshed by periodic Path and Resv messages; otherwise, it will be terminated after a *cleanup timeout* interval. RSVP's default interval for this cleanup timeout is some multiple of the period of the Path and Resv messages. Therefore, if the router misses a single Path or Resv refresh, it will not terminate the session. This kind of tolerance is necessary, since there is no preferential QoS treatment for RSVP messages inherent to the protocol. These messages are sent as best effort unless some provision has been made otherwise, such as DiffServ. While this is generally not necessary, in an environment where the traffic receiving the reservations is extremely critical (such as might be the case on a trading floor or other financial institution), it might be best to err on the side of caution.

NOTE

This is a good time to make mention of the fact that any AVVID, QoS or other network deployment should always be fully tested in a lab environment, and then tested in a noncritical subset of the production network before large-scale deployment is considered. Only after the solution has performed to specifications in both of these environments should it be considered ready for production deployment.

These soft states are dynamic in response to route changes in the network, changes in senders or receivers, or changes in the requested QoS. There is no real difference between the process of initiating a new reservation and refreshing an old one. In both cases, the Path message is built with the previous hop and next hop information, and the Resv statement is adjusted with any new QoS requirements.

NOTE

The refresh interval presents a potential problem when the routing changes across an IP network. If a route change causes any change to the shortest path for an active flow, packets will be forwarded over the new route as best effort until Path and Resv messages are sent along this new path. When this occurs, it is possible that there may not be the necessary resources to complete the RSVP session. In this case, it is up to the application to decide whether to terminate the session or continue best-effort delivery. While Video traffic is somewhat forgiving, this behavior can cause significant problems for voice, therefore, implementing RSVP in this way may not give your traffic the desired results (i.e., good quality) in an unstable network or one with frequent routing changes.

Good implementations of RSVP will issue tear-down messages when the reservation is no longer needed, instead of waiting for the cleanup timeout to remove the session. There are two types of tear-down messages: PathTear and ResvTear. PathTear messages, like Path messages, flow in the downstream direction, whereas ResvTear messages, like Resv messages, flow upstream. In addition to clients issuing immediate requests for tear-downs, routers detecting a session timeout or a loss of resources will send their own tear-down messages to upstream (ResvTear) and downstream (PathTear) neighbors.

What Kind of QoS Can I Request with RSVP?

The types of QoS that can be reserved by RSVP are consistent with the Internet Integrated Services Model. These types are *controlled-load* and *guaranteed-rate*. According to the Intserv definition, controlled-load gives applications service as if they were traversing an unloaded network. Applications requesting controlled-load can expect low latency and a low number of packet drops. These applications are usually considered *tolerant real-time* applications. An example could be an adaptive real-time application like the playback of a recorded conference call. On

Cisco routers, controlled-load services are implemented primarily with weighted random early detection. We will discuss WRED later in this chapter.

RSVP can also request guaranteed-rate services. According to the Intserv definition, guaranteed-rate delivers assured bandwidth with consistent delay. Applications that require this service to function well are usually considered *intolerant real-time* applications. An example would be delay-sensitive applications like VoIP. On Cisco routers, guaranteed-rate services are implemented primarily with weighted fair queuing.

NOTE

Although WFQ can provide guaranteed-rate services to applications, it alone may not be sufficient to assure low latency to delay-sensitive applications such as VoIP during periods of congestion. IOS versions 12.1(3)T and later provide support for LLQ to RSVP.

Reservation Styles and Merging Flows

When a reservation is made, a set of options can be chosen that is collectively called the reservation style. RSVP supports two classes of reservations, *shared* and *distinct*, and two scopes of reservations, *explicit* and *wildcard*. A *shared* reservation is a single reservation made for all packets from multiple upstream senders. A *distinct* reservation establishes a reservation for each sender. For the scope, an *explicit* list can be chosen for the senders, in which each sender is enumerated. The other scope option is to use a *wildcard* that implicitly selects all the senders.

These options give rise to three possible reservation styles (see Table 8.2). The combination of a distinct reservation with a wildcard scope is disallowed and is therefore not defined.

Table 8.2 RSVP Reservation Styles

Scope	Distinct Reservations	Shared Reservations
Explicit	fixed-filter (FF) style	shared-explicit (SE) style
Wildcard	not defined	wildcard-filter (WF) style

These different styles are chosen based on the type of transmitted data that will comprise the reservation.

Wildcard-Filter Style

The combination of a shared reservation and a wildcard sender selection gives the wildcard-filter (WF) style. In this style, a single reservation is made and shared by *all* upstream senders. Reservations can be thought of as a shared pipe whose size is the largest of the resource requests from all receivers for that link, independent of the number of senders.

Shared-Explicit Style

The combination of a shared reservation and an explicit sender list gives rise to the shared-explicit (SE) style. The SE style creates a single reservation shared by a *list* of upstream senders. Both the WF and SE reservation styles are appropriate for data flows that are known not to interfere with each other. An example of this would be an audio conference where it could be assumed that the multiple senders would not typically talk at the same time. It might, however, be wise to make the reservation twice for an audio conference in order to allow for some over-talking, which is a common occurrence, while still retaining QoS.

Fixed-Filter Style

The combination of a distinct reservation and an explicit sender list gives rise to the fixed-filter (FF) style. In this style, a distinct reservation is created for data packets from a particular sender. This reservation is not shared with any other sender. However, if another receiver is added for that sender, the reservation is not doubled, but merged. In an AVVID environment, this kind of style would be appropriate for video signals where the data from each sender are different.

An RSVP *flow descriptor* is the combination of a *flowspec* and a *filterspec*. A *flowspec* is the QoS requested, and the *filterspec* is the set of packets to receive this QoS. When new flows are added to the group of reservations in a node, they will often need to be merged into a common reservation. In the case of multicast traffic, where the same data is going to multiple recipients, the recipients will still make a Resv request. It is up to RSVP to join this request with the active reservations. When this is done, the flows are referred to as *merged*.

The RSVP rules do not allow the merging of distinct and shared reservations, nor the merging of explicit sender selection and wildcard sender selection. As a result, all three styles are mutually incompatible.

Designing & Planning…

Subnetwork Bandwidth Manager

We have seen that RSVP is largely independent of the media it is running on with respect to the QoS mechanisms used to implement a particular reservation. With serial links, WFQ and WRED can be used to provide either a controlled-load or a guaranteed-rate to an application. These mechanisms are not appropriate on a shared medium like Ethernet that has multiple participants competing for the bandwidth. Because of its end-to-end signaling nature, without a common node to keep track of active reservations, a RSVP client on a shared medium would have no way of knowing if there are resources available for the new reservation.

Subnetwork Bandwidth Manager (SBM) was created to implement RSVP on IEEE 802-based networks (Ethernet/Token Ring). SBM acts very much like RSVP. On a shared medium, all SBM-enabled nodes elect a Designated SBM to manage the resources for that network. All RSVP requests by clients on this network are sent to the DSBM for verification. If the resources are available, the request is sent on to the destination address. If the resources are not available, the request is denied.

When using SBM, in order to guarantee that RSVP sessions are not overwhelmed by non-RSVP traffic, you must ensure that all nodes connected to the shared media are RSVP-compliant. This might be difficult to put into practice.

Depending on the topology, SBM will not always be necessary to provide good end-to-end performance for critical applications. Just because part of the journey that a packet takes includes a non-RSVP shared medium such as Ethernet, does not mean that QoS will be impossible. Consider the case of a switched 100BaseTX network connected to a wide area network (WAN) via a T1 on a serial interface of a local router. If it can be assumed that all RSVP requests are destined across the WAN, the bottleneck is clearly the T1. If there are available resources on the T1, it is probable that there are available resources on the Ethernet segment, assuming that non-RSVP flows do not overwhelm the RSVP sessions.

Why Do I Need RSVP on My Network?

RSVP is used primarily to guarantee QoS to real-time applications such as voice and video, which makes it a natural part of any discussion about AVVID solutions.

RSVP-aware clients can make a reservation and be guaranteed a good QoS across the network for the length of the reservation, however long or short.

Because RSVP takes the Intserv approach to QoS, all traffic in the network does not need to be classified in order to give proper QoS to RSVP sessions. On the other hand, for the same reason, a multi-field classification must be performed on each packet at each node in the network to discover if it is part of the RSVP session for which resources have been reserved. This can lead to a consumption of network resources like memory and CPU cycles.

RSVP's open architecture and transparency allow for deployment on many platforms, and even tunneling across non-RSVP aware nodes. Despite this, RSVP has some distinct scaling issues that make it doubtful it will ever be implemented successfully on a very large network, or the Internet for that matter, in its current revision. These advantages and disadvantages, as well as others previously discussed, are summarized here.

Advantages of Using RSVP

- **Admissions Control** RSVP not only provides QoS, but also helps other applications by *not* transmitting when the network is busy.
- **Network Independence/Flexibility** RSVP is not dependent on a particular networking architecture.
- **Interoperability** RSVP works inside existing protocols and with other QoS mechanisms.
- **Distributed** RSVP is a distributed service and therefore has no central point of failure.
- **Transparency** RSVP can tunnel across an RSVP-unaware network.

Disadvantages of Using RSVP

- **Scaling Issues** Multifield classification and statefulness of reservations may consume memory and CPU resources.
- **Route selection and stability** The shortest path may not have available resources, and the active path may go down.
- **Setup time** An application cannot start transmitting until the reservation has been completed.

Using Class-Based Weighted Fair Queuing

Priority Queuing (PQ) and Custom Queuing (CQ) can be used to give certain types of traffic preferential treatment when congestion occurs on a low-speed serial link, and Weighted Fair Queuing (WFQ) automatically detects conversations and attempts to guarantee that no one conversation monopolizes the link. These mechanisms, however, have some scaling limitations. PQ/CQ simply cannot scale to handle links much higher than T1, and the WFQ algorithm runs into problems as traffic increases or if it is stressed by many conversations. Additionally, it does not run on high-speed interfaces such as ATM. Class-based weighted fair queuing (CBWFQ) was developed to overcome these issues and provide a truly scalable QoS solution. CBWFQ carries the WFQ algorithm further by allowing user-defined classes, which allow greater control over traffic queuing and bandwidth allocation. CBWFQ provides the power and ease of configuration of WFQ, along with the flexibility of custom queuing. This advanced queuing mechanism also incorporates weighted random early detection. WRED is not necessary for the operation of CBWFQ but works in conjunction with it to provide more reliable QoS to user-defined classes. We discuss WRED in more detail later in this chapter.

CBWFQ is a very powerful congestion management mechanism and, although it is still being developed to be even more robust and intelligent, its wide platform support and functionality make it an excellent candidate for consideration as part of your end-to-end QoS solution.

How Does CBWFQ Work?

Flow-based WFQ automatically detects flows based on characteristics of the third and fourth layers of the OSI model. Conversations are singled out into flows by source and destination IP address, port number, and IP precedence.

If a packet going out an interface needs to be queued because of congestion, the conversation it is part of is determined, and a weight is assigned based on the characteristic of the flow. Such weights are assigned to ensure that each flow gets its fair share of the bandwidth. The weight assigned also subsequently determines which queue the packet will enter and how it will be serviced.

The limitation of flow-based WFQ is that the flows are automatically determined, and each flow gets a fair share of the bandwidth. This fair share of the bandwidth is determined by the size of the flow and moderated by IP precedence.

Packets with IP precedences set to values other than the default (zero) are placed into queues that are serviced more frequently, based on the level of IP precedence, and thus get a higher overall bandwidth. Specifically, a data stream's weighting is the result of some complex calculations, but the important thing to remember is that weight is a relative number and the lower the weight of a packet, the higher that packet's priority. The weight calculation results in a weight, but the most important thing isn't that number—it's the packet's specific handling. Thus, a data stream with a precedence of 1 is dealt with twice as fast as best-effort traffic. However, even with the action of IP Precedence on WFQ, sometimes a specific bandwidth needs to be guaranteed to a certain type of traffic. CBWFQ fulfills this requirement.

CBWFQ extends WFQ to include user-defined classes. These classes can be determined by protocol, Access Control Lists (ACLs), IP precedence, or input interfaces. Each class has a separate queue, and all packets found to match the criteria for a particular class are assigned to that queue.

Once the matching criteria are set for the classes, you can determine how packets belonging to that class will be handled. It may be tempting to think of classes as having priority over each other, but it is more accurate to think of each class having a certain guaranteed share of the bandwidth. Note that this bandwidth guarantee is not a reservation as with RSVP, which reserves bandwidth during the entire period of the reservation. It is, instead, a guarantee of bandwidth that is active only during periods of congestion. If a class is not using the bandwidth guaranteed to it, other traffic may use it. Similarly, if the class needs more bandwidth than the allocated amount, it may use or borrow some of the free bandwidth available on the circuit.

You can specifically configure the bandwidth and maximum packet limit (or queue depth) of each class. The weight assigned to the class's queue is calculated from the configured bandwidth of that class. As with WFQ, the actual weight of the packet is of little importance for any purpose other than the router's internal operations. What is important is the general concept that classes with high assigned bandwidth get a larger share of the link than classes with a lower assigned bandwidth.

CBWFQ allows the creation of up to 64 individual classes, plus a default class. The number and size of the classes are, of course, based on the bandwidth. By default, the maximum bandwidth that can be allocated to user-defined classes is 75 percent of the link speed. This maximum is set so there is still some bandwidth for Layer 2 overhead, routing traffic (BGP, EIGRP, OSPF, and others), and best-effort traffic. Although not recommended, it is possible to change this maximum for very controlled situations in which you want to give more bandwidth

to user-defined classes. In this case, caution must be exercised to ensure you allow enough remaining bandwidth to support Layer 2 overhead, routing traffic, and best-effort traffic.

Each user-defined class is guaranteed a certain bandwidth, but classes that exceed that bandwidth are not necessarily dropped. Traffic in excess of the class's guaranteed bandwidth may use the free bandwidth on the link. *Free* is defined as the circuit bandwidth minus the portion of the guaranteed bandwidth currently being used by all user-defined classes. Within this free bandwidth, the packets are considered by fair queuing along with other packets, their weight being based on the proportion of the total bandwidth guaranteed to the class. For example, on a T1 circuit, if Class A and Class B were configured with 1000 Kbps and 10 Kbps, respectively, and if both were transmitting over their guaranteed bandwidths, the remaining 534 Kbps (1544–1010) would be shared between the two at a 100:1 ratio.

All packets not falling into one of the defined classes are considered part of the default class (or class-default, as it appears in the router configuration). The default class can be configured to have a set bandwidth like other user-defined classes, or configured to use flow-based WFQ in the remaining bandwidth and treated as best effort. The default configuration of the default class is dependent on the router platform and the IOS revision.

Even though packets that exceed bandwidth guarantees are given WFQ treatment, bandwidth is, of course, not unlimited. When the fair queuing buffers overflow, packets are dropped with tail drop unless WRED has been configured for the class's policy. In the latter case, packets are dropped randomly before buffers totally run out in order to signal the sender to throttle back the transmission speed. This random dropping of packets obviously makes WRED a poor choice for classes containing critical traffic. We will see in a later section how WRED interoperates with CBWFQ.

Why Do I Need CBWFQ on My Network?

You might ask yourself, "Why do I need *any* kind of special queuing?" Packet-based networks drop packets by their very nature. IP network protocols are designed around the inevitability of dropped packets. The question therefore becomes, "If you had a choice, which packets would you prefer to keep and which would you prefer to drop?" This will help determine what type of queuing mechanism you choose.

Designing & Planning…

The Battle of the Internet Protocols

Protocols can be categorized as either *congestion notification responsive* or *congestion notification unresponsive*. The *slow start* algorithm characterizes the Transmission Control Protocol (TCP) as being responsive to congestion situations since when a TCP flow fails to get an acknowledgement that a packet was received, it throttles back its send rate and then slowly ramps up.

On the other hand, the User Datagram Protocol (UDP) is unresponsive to congestion notification. Unless there are acknowledgements at a higher layer, a UDP stream will continue to transmit at the same rate despite packet drops. If the traffic is a mixture of TCP and UDP, then TCP is polite and UDP is usually the spoiler. The unresponsiveness of UDP applications can be the detriment of not only other *impolite* UDP streams but also well-behaved TCP sessions.

WFQ is on by default on low-speed serial interfaces for good reason. It works well to overcome the limitations of first in/first out (FIFO) queuing by not allowing large flows to dominate smaller, interactive flows, and it is easy to implement. However, even with the extension of the weighting model by using IP precedence, flow-based fair queuing is still just that—fair. There are times when the fair slice of the bandwidth pie is less than you require for certain applications, or when you require more granular control over the QoS provided to your traffic.

With CBWFQ, you can leverage the DiffServ model to divide all your traffic into distinct classes to which CBWFQ can subsequently give specialized bandwidth guarantees. The typical application of this is to mark traffic at the edge with IP precedence, and then let mechanisms like CBWFQ give differential treatment throughout the entire network according to the service levels defined. By placing important applications into a class to which CBWFQ can give a guaranteed bandwidth, you have effectively prevented other applications from stealing bandwidth from those critical applications. Let us examine a couple of illustrative cases.

NOTE

Advanced queuing mechanisms (basically, anything except FIFO) work to schedule which of the packets waiting in queue will be next to go out the interface. Thus, advanced queuing mechanisms really do not come into play unless there is congestion. If there are no packets waiting in queue, then as soon as a packet comes into the router, it goes directly out of the interface, and the queuing works essentially the same as FIFO. Therefore, CBWFQ does not kick in until congestion starts.

Case Study: Using a SQL Application on a Slow WAN Link

Imagine that Company A uses a SQL application for centralized inventory. It was originally used only at the corporate headquarters; however, it has now become critical to the core business, and its use has been extended to remote sites. Unfortunately, because it was developed in a LAN environment, it does not respond well to delays and packet loss. Assume that it needs 50 Kbps to function adequately, and that all the remote sites are connected with 256 Kbps serial links. In the absence of other traffic, the application functions perfectly. However, at peak times during the day, other applications such as bulk transfers from FTP, Telnet sessions to the corporate mainframe, Web browsing, and messaging are periodically filling the link to capacity. With WFQ enabled, some SQL packets may be dropped in a congestion situation because of the competing conversations. Remember that all traffic gets its fair share of the bandwidth and its fair share of packet drops. The drops would cause TCP retransmissions, which could slow down the SQL application considerably. Because of the SQL application's interactive nature, the user's productivity drops, and he or she comes to you requesting an upgrade of the link speed. A circuit upgrade might sound like a good idea if we could get the project funding. However, if we did this, we might quickly find out that even if we doubled the circuit speed, the company's critical application might still not achieve the performance it requires. IP networks work in bursts, and even the largest pipes can momentarily become saturated.

One solution would be to configure a class for the SQL application. The SQL traffic could be classified by the TCP port number of incoming packets. By applying a policy to the output of the serial interface allocating 50 Kbps to this

class, we could guarantee that even during the busiest part of the day, this application would be given the amount of bandwidth needed for good performance. In addition, all other traffic could be configured to function under flow-based WFQ so all conversations would have fair access to the remaining bandwidth.

In effect, we have carved out a slice of the serial bandwidth for the SQL application but also allowed it to use more than this amount, although its use above 50 Kbps would not be guaranteed. In addition, other applications can use the reserved 50 Kbps when SQL is not using it. Remember, CBWFQ does not function unless there is congestion.

Case Study: Total Traffic Classification (CBWFQ in a DiffServ Model)

In the previous case study, we saw that we could effectively guarantee a certain amount of bandwidth to a mission-critical application. But what if there were many other applications that needed minimum bandwidth guarantees? (We will address voice, and other truly latency-sensitive types of traffic in just a minute.) We may need more granular control over how our applications behave under WFQ. CBWFQ allows us to configure up to 64 distinct classes. However, we probably would not want to put each application into a separate class. Not only would we be limited in the amount of bandwidth we could allocate to the class (the sum of all bandwidth cannot exceed the link speed), but it could also be confusing having this many classes.

A best-practice approach would be to define just a few of the classes, and categorize all applications into these classes based on expected bandwidth utilization and the application's tolerance of dropped packets. With this approach, applications would be sharing bandwidth with others within the same class, but a degree of granularity is added in addition to WFQ that would be adequate for most networks.

The IP CoS header allows us to enumerate packets into eight levels of IP precedence, two of them being reserved for network applications, leaving six levels for user applications. We can map these IP precedence levels directly into our network classes of service. Using a precious metal analogy, we would have six classes of service, as shown in Table 8.3.

Table 8.3 An Example of a Class of Service Mapping

Class of Service	IP Precedence
Platinum (Typically Voice Traffic)	5
Gold	4
Silver	3
Bronze	2
Iron	1
Best-Effort (default)	0

In this example, we can realize the economy of using CBWFQ within the DiffServ model. Using packet classification at the edge of the network to mark IP precedence, we have effectively divided all our applications into five classes of service, plus a default class. Except for the edge devices, no other classification may be necessary to place a packet into the proper queue as it traverses the network. By marking applications at the edge and allowing internal routers to queue packets according to these classes, we not only assure consistent QoS for that application across the entire network, but we also reduce the resource load on both the routers and the network administrator. The routers do not have to process lengthy ACLs at every hop, and the administrators have to worry about classification only at the edge of the network. Additionally, it is at these edge devices that packet rates are the smallest, and processor utilization according to packet marking is manageable. To classify packets at the hub site where many circuits are being aggregated might be too much for the router to handle.

NOTE

Remember that QoS is never a substitute for bandwidth. On the other hand, even a gigabit link can drop packets if the queues fill up. Congestion management rations the limited bandwidth to the most important applications, or in the case of CBWFQ, ensures that certain applications get at least the percentage of total bandwidth allocated. The important point here is that QoS mechanisms will help prioritize traffic on a congested link (and drop the least important traffic first and most often) but, at some point, a link may become so congested that packet drops reach an unacceptable level. When this point is reached, a bandwidth upgrade is in order.

RSVP in Conjunction with CBWFQ

CBWFQ and RSVP can be configured on the same interface. There is, in general, no specific interaction between the two. They are configured as if the other mechanism were not present. However, because RSVP reserves bandwidth for its clients and CBWFQ guarantees bandwidth for its classes, it is possible to configure the router to guarantee bandwidth to each of them in such a way that the total guaranteed bandwidth exceeds the circuit speed.

This constitutes a potential problem. In a congestion situation, if you have promised the majority of the circuit bandwidth to two mechanisms separately, which one will succeed in getting the bandwidth it needs? You cannot promise three-quarters of the bandwidth to CBWFQ and half the bandwidth to RSVP and expect that they would both have sufficient bandwidth in a congestion situation. In practice, if you need to guarantee bandwidth to classes as well as to RSVP sessions, you would avoid an overlapping bandwidth guarantee like this. Still, there is nothing in the IOS code to prevent you from making this configuration.

So, what exactly does happen if you over-subscribe the guaranteed bandwidth by promising it to both RSVP and CBWFQ? Because of the WFQ implementation in the routers, RSVP wins out in the end, taking as much bandwidth as it needs from all other classes equally.

Using Low Latency Queuing

The previous section demonstrated that CBWFQ can give bandwidth guarantees to different classes of traffic. Although CBWFQ can provide these bandwidth guarantees, low latency transmission may not be provided to packets in congestion situations, since all packets are transmitted fairly based on their weight. This can cause problems for applications like VoIP that are sensitive to delays, especially variations in delays. Variation in the delay time between individual packets that make up a voice stream is usually referred to as *jitter*. Although most voice applications can tolerate a certain amount of delay, jitter can cause choppiness in voice transmissions and quickly degrade overall voice quality. Low Latency Queuing (LLQ) extends CBWFQ to include the option of creating a strict priority queue. Strict priority queuing delivers low latency transmission to constant bit rate (CBR) applications such as voice. Due to the nature of LLQ, it is not recommended that you configure anything other than voice traffic to be placed in the priority queue, as this can cause serious problems for your voice traffic.

How Does LLQ Work?

Once you know how CBWFQ works, LLQ is easy to understand. LLQ creates a strict priority queue that you might imagine as resting on top of all other queues. This priority queue is emptied before any other queue is serviced. A strict priority queue is often referred to as an exhaustive queue, since packets continue to be removed from the queue and transmitted until it is empty. Only after the strict priority queue is totally empty are the other queues serviced in the order determined by whatever weighting has been configured by the CBWFQ bandwidth statements. If you're thinking this sounds an awful lot like the much older QoS technique, simply called "Priority Queuing," you're absolutely correct. Think of LLQ as a hybrid, formed from the union of CBWFQ and Priority Queuing.

NOTE

When LLQ was first created, it was referred to as PQCBWFQ, or priority queuing with class-based weighted fair queuing. Although this lengthy acronym was appropriate because it clearly described the combined functionality of PQ with CBWFQ, it has been changed in most documentation to simply LLQ.

If packets come into the priority queue while another queue is being serviced, the packets waiting in the priority queue will be the very next packets sent out the interface after the current packet has been transmitted. In this way, the delay between packets sent from the priority queue is minimized, and low latency service is delivered. The maximum time between priority packets arriving at the far end would occur in the case in which a packet arrives in the previously empty priority queue as soon as the router starts to transmit a large packet. The largest possible packet is referred to as the maximum transmission unit (MTU), which is 1500 bytes on Ethernet. The priority packet will have to wait for the nonpriority packet to finish transmitting. Thus, the longest delay possible between arriving priority packets is limited to the serialization time of the MTU plus the serialization time of the priority packet itself. The serialization time is calculated by dividing the size of the packet by the link speed (packet size/link speed). We discuss the implications of serialization delay and how to overcome it in more detail in a later section on Link Fragmentation and Interleaving (LFI).

Classifying Priority Traffic

The traffic placed into the priority queue under LLQ is determined by the same criteria available to any other user-defined class under CBWFQ. Specifically, these criteria include protocol, ACLs, IP precedence, and input interface. As mentioned in Table 8.3, IP Precedence 5 is generally used for Voice traffic. The most common way to determine which traffic goes into the LLQ is to match the IP Precedence, since many devices (including Cisco IP Phones) automatically set the IP Precedence to 5 on Voice traffic.

Allocating Bandwidth

Bandwidth is allocated to the priority class a little differently than to other user-defined classes. Instead of specifying the guaranteed bandwidth of the class with the **bandwidth** command, the **priority** command is used. This gives a priority class that will deliver LLQ to all traffic falling under this classification. There is a particular distinction between how traffic metering is handled with the priority class as opposed to other user-defined classes. Unlike normal classes, with the priority class under congestion situations, bandwidth in excess of the limit configured with the **priority** command *is always dropped* on the 7200 Series and lower Cisco platforms. This is to prevent the priority queue from starving other traffic, both user-defined classes and other important traffic like network routing updates. However, in noncongestion situations, the bandwidth allocated to the priority class may be exceeded.

It is important that you limit the bandwidth allocated to the priority class to a reasonable value. If you configure too much of your traffic as priority traffic, then it really is not priority at all. On an airplane, if everyone flies first class, can you really call it first class? Additionally, it is strongly recommended that packets classified into the priority class be limited to voice traffic alone, as mentioned earlier. Voice streams are made of small packets of constant bit rate that are well behaved by nature. By classifying applications into the priority class that are prone to bursts or comprised of large packets, you essentially destroy the low latency provided to the small-packet CBR voice traffic also waiting in the priority queue.

The fact that bandwidth of the priority class under congestion situations creates a "hard upper limit" to voice traffic should not cause insurmountable problems. Voice planners are accustomed to providing for an exact number of voice calls on traditional voice networks. The same can be done on VoIP networks by multiplying the bandwidth of each voice call (determined by the CODEC) by the number of simultaneous calls in order to get the bandwidth necessary. It is

important to note that a call admission control process for the voice calls is required. This guarantees that the number of calls supported by the bandwidth provisioned by the **priority** command is not exceeded. Exceeding this bandwidth would potentially lead to poor voice performance for *all* voice callers. Here is an example.

Consider that a remote site needs up to 24 simultaneous voice calls connected to the main hub site. The remote site is connected via a T1 serial link. When the G.729 CODEC is used with cRTP, you can expect each call to use a maximum of 12 Kbps. This gives a provision of 288 Kbps for all 24 calls. This bandwidth is configured for the priority class with the **priority** command. In an uncongested situation, more than 24 calls could be completed and still have good quality. However, if congestion occurs in this overloaded call state, even for a moment, packets will be dropped from the priority queue. Since it can be assumed that the packets from the individual voice calls are interleaved with each other, some drops will occur across all connected voice calls, resulting in poor performance for everyone. To avoid this, some kind of admission control system is necessary to assure that no more than 24 calls are ever connected. This can be accomplished in a number of ways, including using gatekeeper technology available on the Cisco Call Manager, the Cisco AS5300, and Cisco 3640 routers (IOS 12.1(1)), or by limiting the number of active voice ports on communicating gateways. In either case, it would be preferable for a caller to get a busy signal indicating that the call could not be completed or to have exceeding calls re-routed to the PSTN, rather than the quality of all connected callers being affected.

Limitations and Caveats

A notable difference between the priority class and other user-defined classes under CBWFQ is that WRED is not available in the priority class. LLQ is to be used for CBR services, especially VoIP. Voice traffic is UDP-based and therefore not adaptive to the early packet drop characteristic of WRED. If a packet is dropped from a UDP stream, UDP will not react to this by reducing the send rate. Because WRED would be ineffective, configuration of this feature for a priority class using the **random-detect** command is disallowed. Besides, would you *really* want to randomly drop your voice traffic, anyway? Not likely.

Why Do I Need LLQ on My Network?

You should consider using LLQ if you need to provide good QoS to delay- and jitter-sensitive applications like VoIP. Because LLQ is an extension of CBWFQ, it

complements network designs that are already using CBWFQ to give differential services to classes of applications. You have only to configure another class and designate it as "priority" with an appropriate bandwidth limitation to give low latency service to your real-time applications.

Because LLQ is an extension of CBWFQ, you also have access to all the matching criteria that is provided normally to CBWFQ. This is in contrast to RTP priority queuing, which limits match criteria to a UDP port range. Since one of these matching criteria is IP precedence, the DiffServ model can be leveraged to use packet marking at edge devices and allow CBWFQ with LLQ to give low latency service to designated packets without long ACLs. This speeds up packet processing time and overall performance. LLQ is also more flexible than RTP priority queuing in that it can be enabled on ATM virtual circuits (VCs) to allow timely de-queuing of delay-sensitive traffic into ATM networks.

Finally, the hard limit of the bandwidth for priority classes acts as a sort of traffic cop that prevents LLQ from starving other traffic classes of bandwidth in congested situations.

Using Weighted Random Early Detection

Random Early Detection (RED) can be used as a congestion avoidance mechanism to prevent congestion problems at bandwidth bottlenecks on networks. Weighted Random Early Detection (WRED) is the Cisco implementation of RED that combines the RED algorithm with weighting determined by IP precedence levels. This effectively gives higher precedence traffic lower drop rates and thus priority over lower precedence traffic in the network.

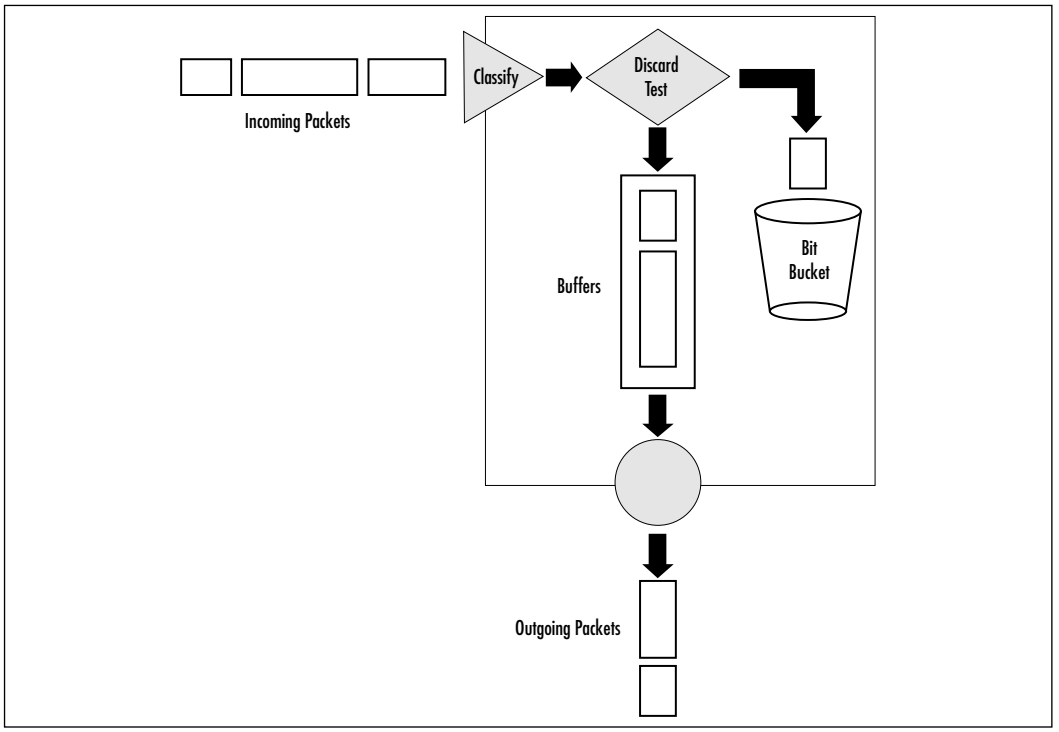
How Does WRED Work?

RED works on the basis of active queue management, and addresses the shortcomings of tail drop. A RED-enabled router signals congestion to TCP senders by dropping packets before the router is actually out of buffer space. Compliant TCP senders detecting the dropped packets will throttle back the send rate using the TCP slow start algorithm. RED drops arriving packets randomly so the probability of a particular flow having packets dropped is in proportion to the flow's share of the bandwidth. Thus, flows using more bandwidth have a greater chance of dropped packets than flows using small amounts of the overall bandwidth.

RED operates by monitoring the buffer level and discarding packets probabilistically (see Figure 8.3) based on minimum and maximum threshold values.

Below the minimum threshold, no packets are dropped; above the maximum threshold, all packets are dropped. When the buffer is between these two thresholds, the drop rate is calculated as a function of the average queue size. The average queue size is a running average over time. How responsive this average is to changes is reflected in the configurable weighting average (discussed later). Because of the randomness in which packets are dropped, packets across all flows are dropped at different times, thus preventing the phenomenon of *global synchronization* commonly associated with tail drop.

Figure 8.3 Weighted Random Early Detection



WRED and IP Precedence

WRED is the Cisco implementation of RED that combines the capabilities of the RED algorithm with IP precedence to provide lower drop rates for higher priority, or higher precedence, packets. The router attributing different minimum and maximum threshold levels to each precedence level accomplishes this. By default, the minimum threshold in packets for IP precedence level 0 is one half the maximum. The values for the remaining precedences fall between half the

maximum threshold and the maximum threshold at evenly spaced intervals. Table 8.4 shows the default values for both WRED and Distributed WRED (DWRED), which is available on the Versatile Interface Processors (VIP)-based Route/Switch Processor (RSP) platform. See the discussion later in this chapter on "Running in Distributed Mode."

NOTE

Although WRED gives lower drop probabilities to higher IP precedence values, it can be configured to change the weighting of each precedence level, or even to ignore precedence altogether, and thus function as normal RED. By using the **random-detect precedence** command, you can set the minimum and maximum threshold levels to something other than the default values shown in Table 8.3. By making all the thresholds the same, you essentially make WRED function as normal RED.

Table 8.4 Default WRED and DWRED Threshold Values

IP Precedence	WRED Threshold Values		DWRED Threshold Values	
	Minimum	Maximum	Minimum	Maximum
0	20	40	95	190
1	22	40	106	190
2	24	40	117	190
3	26	40	128	190
4	28	40	139	190
5	31	40	150	190
6	33	40	161	190
7	35	40	172	190
RSVP	37	40	N/A	N/A

WRED and RSVP

WRED is the primary QoS mechanism responsible for providing *controlled-load* service to RSVP sessions. Remember from our RSVP discussion that Intserv defines *controlled-load* service as service across the network as if it were unloaded. By WRED keeping a link in a noncongested state by detecting impending congestion

situations and preemptively dropping traffic, WRED can effectively grant services to RSVP flows *as if the network were unloaded*.

WRED Algorithm

The basic RED algorithm uses a calculated average queue size to determine when to drop packets and with what probability. This average is based on the previous average and the current queue size. It therefore can be considered a moving average with the following formula:

$$average = (old\_average * (1-2^{-n})) + (current\_queue\_size * 2^{-n})$$

In this equation, *n* is the exponential weighting constant that affects how rapidly the average changes with respect to the current queue size. By changing this constant, WRED can be configured to be more or less adaptive to bursts in traffic. Cisco recommends using the default value of 9, but you can change this by using the **random-detect exponential-weighting-constant** command. Valid ranges are between 1 and 16. Higher values will make the moving average slower, which smoothes out the peaks and lows in queue length at the expense of not reacting to congestion fast enough. Lower values will make WRED more adaptive but the possibly exists that WRED may overreact to temporary traffic bursts and drop traffic unnecessarily.

Why Do I Need WRED on My Network?

WRED makes early detection of congestion possible and provides differentiated services for multiple classes of traffic. Like basic RED, it also protects against *global synchronization* as long as flows are compliant to congestion notification, like TCP. For these reasons, WRED is useful on any output interface where you expect congestion to occur. However, it is most effective in backbone networks that aggregate many paths from remote sites. If the routers at the edge are marking traffic into classes with IP precedence, WRED can act more intelligently in the backbone with its drop decisions.

WRED was primarily designed for use in IP networks dominated by TCP. You should use caution when evaluating WRED if your network has a large amount of UDP traffic, because UDP traffic simply does not respond to packet drop in a manner that would allow WRED to provide any relief. With TCP, dropped packets indicate congestion, so the packet source will reduce its transmission rate. With other protocols, like UDP, it is left to higher layer protocols, such as the application itself, to respond to dropped packets by slowing down transmission. With

UDP, this usually does not happen. When packets are dropped from a UDP transmission, the source may continue to send packets at the same rate. Thus, dropping packets does not decrease congestion, and WRED is ineffective. Making sure that adaptive flows get their fair share of bandwidth in comparison to nonadaptive flows may be possible using flow-based RED (see "Flow-Based Random Early Detection" sidebar).

Additionally if your network is not strictly IP, you may not gain the benefit of the IP precedence weighting of WRED. WRED treats non-IP traffic as precedence 0, the lowest precedence. Therefore, non-IP traffic will be lumped into a single bucket and is more likely to be dropped than IP traffic. This may cause problems if most of your important traffic is something other than IP. The case for QoS may encourage you to advocate transforming your network into a strictly IP network or, at least, tunneling non-IP traffic through the portions of your network where QoS is required.

Designing & Planning…

Flow-Based Random Early Detection

Flow-Based RED (FRED) is an extension to WRED that ensures no single flow can monopolize all the buffer resources at the output interface queue. With normal WRED, a packet dropped from a TCP source causes the source to reduce its transmission, whereas a packet dropped from a noncompliant source, like UDP, does not. This may have the end effect of the *polite* flows being drowned out by the *impolite* flows. Flow-based RED prevents this by maintaining minimal information about the buffer occupancy of each flow. In this way, when a flow exceeds its share of the output buffer, it is dropped. This is in contrast to the more random buffer drops of normal WRED. This feature first became available in IOS version 12.0(3)T.

Using Generic Traffic Shaping and Frame Relay Traffic Shaping

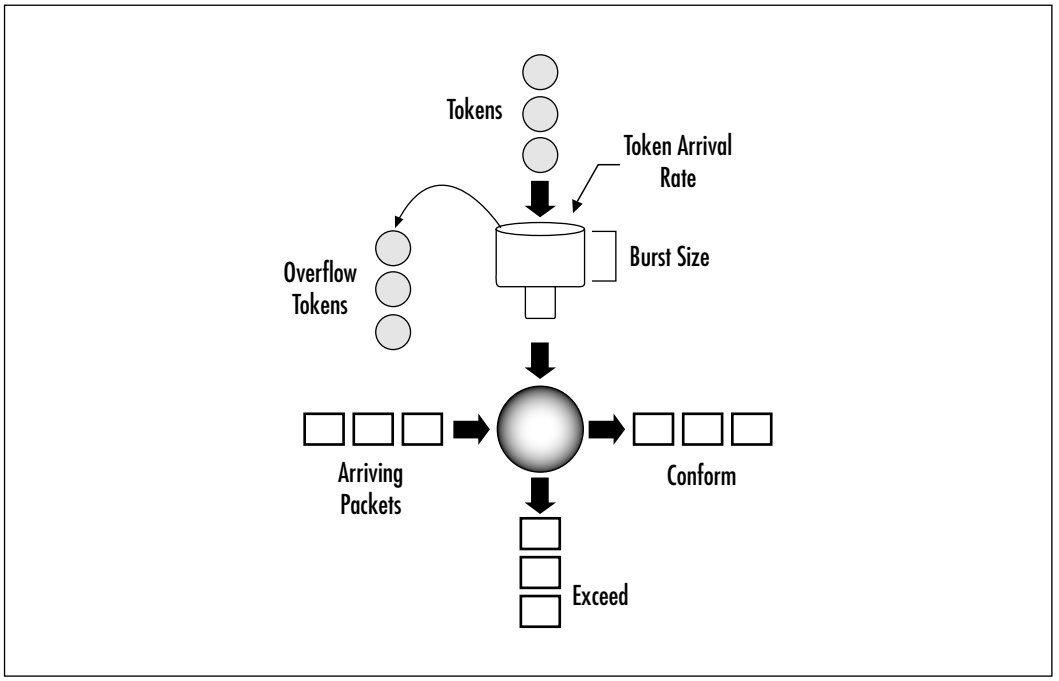
Traffic shaping is a mechanism that restricts traffic going out an interface to a particular speed and, at the same time, attempts to buffer bursts in excess of this

maximum speed. Traffic shaping thereby acts to smooth out or shape traffic into a stream that conforms to downstream requirements. Cisco offers two traffic shaping features, namely, Generic Traffic Shaping (GTS) and Frame Relay Traffic Shaping (FRTS). These two features are more similar than they are different. To understand traffic shaping in general, let us first look at the fundamental algorithm behind it.

Token Bucket

Both GTS and FRTS use a construct called a token bucket to rate-limit traffic. A token bucket should be thought of as being filled with tokens, not packets (imagine tokens as permissions for a specific number of bits to be transmitted to the network). The token bucket is also commonly referred to as a *credit manager* that gives credits to traffic to be used for transmission. Before a packet is sent out the interface, a certain number of tokens need to be removed from the bucket. Tokens fill the token bucket at a constant rate, and the bucket is a certain size. After the bucket is full, newly arriving tokens are discarded. If the bucket is empty, an incoming packet has to wait for enough tokens to fill the bucket before it can be transmitted. Thus, with the token bucket analogy, the burst size is roughly proportional to the size of the bucket. A depiction of a token bucket is shown in Figure 8.4.

Figure 8.4 Token Bucket Algorithm



There are three primary variables associated with token bucket traffic shaping: burst size, mean rate, and time interval.

- **Mean rate** Specifies how much data can be sent on average. This is also called the committed information rate (CIR).
- **Burst size** Specifies how much data can be sent over a single time interval without causing scheduling problems. This is also called the Committed Burst size.
- **Time interval** This is the time quantum for a single burst. It is also called the measurement interval.

The burst size is the amount of data that can be sent with the token bucket over a single time interval. The mean rate is the burst size divided by the time interval. Therefore, when a token bucket is regulating an output interface, its rate over an interval of time cannot exceed the mean rate. However, within that interval, the bit rate may be arbitrarily fast. In this way, large data flows are regulated down to what the network can actually handle, and momentary bursts are smoothed out by buffering, rather than being dropped.

How Does GTS Work?

GTS acts to limit packet rates sent out an interface to a mean rate, while allowing for buffering of momentary bursts. With GTS parameters configured to match the network architecture, downstream congestion can be avoided, eliminating bottlenecks in topologies with data-rate mismatches. GTS has the following characteristics:

- Rate enforcement on a per interface, or subinterface, basis—the mean rate can be set to match the circuit CIR or some other value.
- Traffic selection using ACLs.
- GTS works on many Layer 2 interface types, including Frame Relay, ATM, Switched Multimegabit Data Service (SMDS), and Ethernet.
- It supports backwards explicit congestion notification (BECN) messages for bandwidth throttling.
- It supports WFQ per subinterface.

GTS works with the token bucket algorithm in the following way. When packets arrive at the router, an interrupt occurs. If the queue is empty, GTS consults the credit manager (token bucket) to see if there is enough credit to send

the packet. If there is not, the packet is sent to the queue configured, in this case, WFQ. If there is credit available, the packet is sent to the output interface, and the associated credit value is deducted from the token bucket. Queued packets are serviced at regular time intervals. The credit manager is checked at each time interval to determine if there is enough credit to transmit the next packet waiting in queue. If there is, the packet is sent to the output interface, and the VC is charged the appropriate number of credits.

Why Do I Need GTS on My Network?

Many times a situation exists in which a carrier provides a circuit with a CIR less than the access rate of the physical interface. For example, a Frame Relay service may be provisioned with a 1544 Kbps CIR, but the circuit is delivered on an E1 (2048 Kbps) interface. In the absence of traffic shaping, the router will send up to a rate of 2048 Kbps. This may cause problems, since the traffic in excess of the CIR could be dropped in the Frame Relay network. In this situation, you may get considerably more throughput than the CIR at times, but you are at the mercy of the Frame Relay network. During times when the network is not busy, you may get all your traffic through, but during congested times, many of your packets may be dropped. You may think that any amount of bandwidth over the CIR is a bonus, but when packets like TCP are dropped in large quantities, the retransmission can cause not only increased congestion, but *global synchronization* as well. Additionally, if you are transmitting real-time data, any dropped packets will immediately degrade performance. Depending on your network applications, it may be better to take the more conservative approach by using traffic shaping and sleep soundly knowing you have a reliable service.

Although GTS is available on a variety of interfaces, it may not be that useful in light of other QoS mechanisms and modern technologies. For example, you would rarely want to limit traffic rates on a shared, private medium such as Ethernet, especially if it was switched Ethernet. Also, in the case of ATM, if a variable bit rate (VBR) service was ordered, the carrier would most likely tell you the sustainable cell rate (SCR), peak cell rate (PCR), and maximum burst size (MBS). By configuring an ATM VBR service on the router with these parameters, you have already enabled traffic shaping. Adding GTS on top of this would be redundant. Finally, for Frame Relay circuits, FRTS, not surprisingly, has features that are more suited to this medium.

Designing & Planning…

How Do FECNs and BECNs Work?

Forward explicit congestion notification (FECN) and backwards explicit congestion notification (BECN) are used in networks by intermediary nodes to inform other nodes of congestion that was experienced as a packet traveled across the network. In Frame Relay, setting a specific bit in a normal Frame Relay packet indicates a FECN or BECN. Here's how it works.

If device A is sending data to device B across a Frame Relay infrastructure, and one of the intermediary Frame Relay switches encounters congestion (congestion being full buffers), an over-subscribed port, overloaded resources, and so forth, it will set the BECN bit on packets being returned to the sending device (A), and the FECN bit on the packets being sent to the receiving device (B). This has the effect of informing the sending router to slow down and apply flow control, such as traffic shaping, and informing the receiving device that the flow is congested and that upper layer protocols should expect some delays.

How Does FRTS Work?

FRTS works essentially the same as GTS. It uses a token bucket, or credit manager, algorithm to service the main queuing mechanism and send packets out the interface. It also is commonly used to overcome data-rate mismatches. The most frequently seen instance of this is in a hub and spoke environment where the head-end (hub) has a large amount of bandwidth (perhaps a T1) and the spokes have much smaller amounts of bandwidth (perhaps, 128k each). In this case, if a single remote site is being sent 200k of traffic, the remote site will have a completely saturated line but, because of the speed mismatch, the head-end router's interface will not see congestion. Recall that queuing mechanisms will only kick in when there is congestion, so we need a mechanism to create congestion at the head-end. FRTS does have some unique characteristics that we should explore before proceeding:

- Enhanced queuing support on a per VC basis—both PQ and CQ are available
- Traffic selection using ACLs

- Rate enforcement on a per VC basis—the mean rate can be set to match CIR or some other value
- FRTS supports both BECN and Cisco Foresight congestion notification on a per VC basis

Notice that WFQ is not available (see the following note for the IOS release in which it is available), but PQ and CQ are configurable on a per VC basis. This means that you are not limited to one queuing mechanism for the whole interface, but you can pick the queuing method that suits each VC the best. Additionally, by using ACLs, you can direct traffic to separate VCs, creating a virtual time-division multiplexing (TDM) network. This method may not make the most efficient use of your purchased bandwidth if you pay by CIR, since if there is no incoming traffic for a particular traffic type, the associated VC will be basically empty.

Another approach that would make more efficient use of bandwidth would be to divide your traffic into classes on a single VC. For example, suppose DECnet was a critical application across your Frame Relay network. Using PQ, you could classify all your DECnet traffic into the high priority queue, while classifying other traffic into lower ones. Since all the packets in the high priority queue would be serviced before the lower priority queues, you would ensure that DECnet packets would not be delayed unduly by other traffic.

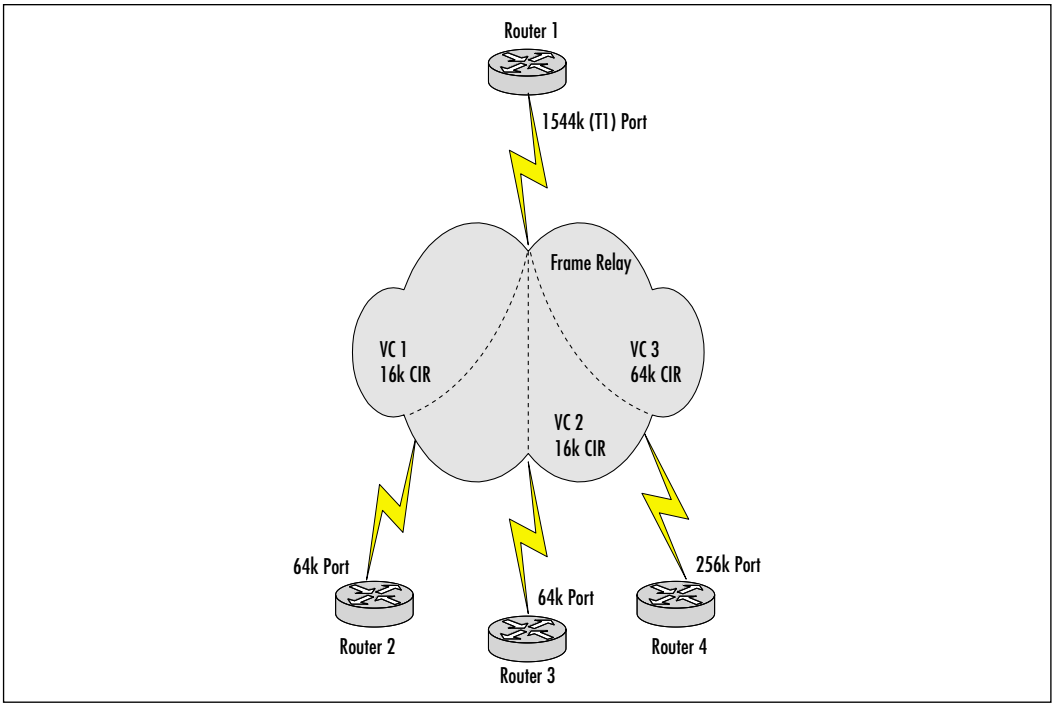
Still another approach would be to divide your traffic into classes and use CQ to give each a guaranteed bandwidth percentage. This has the benefit over multiple VCs and the virtual TDM network of allowing a class's reserved bandwidth to be used by other classes when available.

Why Do I Need FRTS on My Network?

The most common use for FRTS is to overcome data-rate mismatches. Earlier in our discussion of GTS, we saw that sometimes the allowed data rate for the network is lower than the port speed that the interface is delivered on. Beyond this data-rate mismatch that occurs at the router/switch interface, there is a situation that is almost as common that occurs when a site acts as the central hub that terminates many other Frame Relay connections. Consider the example shown in Figure 8.5.

In this example, we see that Router 1 is connected to a Frame Relay switch network (shown as the cloud) via a T1 interface with three virtual circuits. Routers 2 and 3 are connected to different parts of the network, each with a 64 Kbps port speed and a CIR of 16 Kbps. Router 4 is connected with a port speed of 256 Kbps and a 64 Kbps CIR.

Figure 8.5 Traffic Shaping on a Frame Relay Network



With this configuration, we have a separate virtual circuit going to each remote site in a point-to-multipoint configuration. Because of the unequal data rates, without traffic shaping, it is possible that the traffic flowing out of Router 1 might overload any one of the three other routers. Traffic could potentially travel across the majority of the Frame Relay network, only to be dropped at the egress of the network, right before the remote router. This does not make very efficient use of the network. You might consider simply lowering the port speed of the hub router to 64 Kbps to prevent this; however, not only would Router 4 then have the potential to overwhelm the hub router, but if Routers 2, 3, and 4 all transmitted at their port speed simultaneously (64 + 64 + 256 = 384), they definitely would.

FRTS can solve this problem. What is typically done is enable FRTS at the hub location and set the *FRTS CIR* parameter (not the carrier CIR) equal to the port speed at the far end of the VC. Thus, for the first VC from Router 1 to Router 2, we would have a CIR set to 64 Kbps. The same configuration would apply to the second VC. We would set the CIR of the third VC to 256 Kbps. This overcomes the data-rate mismatch, the traffic becomes well-behaved, and unnecessary packet

drops are eliminated. Enabling FRTS on the remote ends might be helpful if you wanted the routers to heed BECNs and throttle down when the network is congested, but by enabling FRTS on the hub site alone, we have eliminated the data-rate mismatch problem. FRTS will heed congestion notifications from both BECN and Cisco Foresight messages.

Configuring & Implementing…

Configuring LLQ for Frame Relay

In earlier releases of IOS, if you wanted to transmit voice over Frame Relay and ensure low latency, you might have considered the combination of RTP priority, PQ or CQ, and FRF.12 for link fragmentation and interleaving (see the section on RTP priority and LFI for more information). However, this combination has been superceded in later releases of IOS code (12.1(2)T or later, to be exact) by the more flexible feature, LLQ for frame relay. The concept and configuration are very similar to general CBWFQ with LLQ covered earlier in this chapter, but a very generic configuration example might be as follows:

```
!
 class-map voice          # We create a class for our Voice
  match access-group 101 # Which we define as packets
                         # matching access-list 101
 !
 class-map video    # We create a class for Video
  match ip prec 4   # Which we define as packets with IP
                    # Precedence of 4
 !
 class-map control   # We create a class for session
                     # control traffic
  match ip prec 3   # Which we define as packets with IP
                    # Precedence of 3
      # There is an implied "class-default"
      # All other packets go go into class-default
```

Continued

```
!
access-list 101 permit udp any any range 16384 20000
```

The following commands create and define a policy map called mypolicy:

```
!
policy-map mypolicy  # We create the policy-map "mypolicy"
 class voice         # We define the class for voice
  priority 16        # The "priority" keyword defines LLQ and
                     # guarantees 16k
 class video
  bandwidth 32  # The "bandwidth" keyword is a bandwidth
                # guarantee (not LLQ)
  random-detect # Turns on WRED within this class
 class control
  bandwidth 16
 class class-default
  fair-queue 64    # Packets in this class will be handled
                   # with regular WFQ
  queue-limit 20   # There can be no more than 20 packets
                   # in this queue
```

The following commands make sure you have Frame Relay traffic shaping at the serial interface. Enable Frame Relay fragmentation and attach the policy map to DLCI 100:

```
!
interface Serial1/0
  frame-relay traffic-shaping # Turns on FRTS on the main
                              # interface
!
interface Serial1/0.1 point-to-point
 frame-relay interface-dlci 100
   class fragment # Assigns class "fragment" to this DLCI
!
```

Continued

```
map-class frame-relay fragment # Defines the class "fragment"
 frame-relay cir 64000
 frame-relay mincir 64000      # CIR and MINCIR are both 64k
 frame-relay bc 640
 frame-relay fragment 50
 service-policy output mypolicy
  # Finally, we assign "mypolicy" to this class
  # which is applied to the DLCI.
```

Running in Distributed Mode

The Cisco 7500 Series is Cisco's high-performance distributed LAN/WAN services router. It follows its predecessor, the 7000 Series, which has been discontinued and will not support IOS revisions above version 11.2. The 7500 Series has architecture quite different from other Cisco router platforms. It is comprised of one Route/Switch Processor and multiple Versatile Interface Processors. Each VIP not only provides modular Port Adapter functionality that supports a wide range of interfaces, but also effectively offloads many tasks from the main RSP. This leads to scalability and an easy network upgrade path. When more capacity is required, additional VIPs can be installed. There are a few types of VIPs that differ in their processing power. In general, higher capacity circuits require faster processors. The true scalability of this platform is realized only when the VIP takes on tasks normally run by the RSP. When these services are run on the individual VIP, the service is said to be running in *distributed mode.* Let's look at some of the features supported in distributed mode.

Features Supported in Distributed Mode

There are many services that can run in distributed mode, including:

- **Basic Switching** Cisco Express Forwarding, IP fragmentation, Fast EtherChannel
- **VPN** IP Security (IPSec), generic routing encapsulation (GRE) tunnels
- **QoS** Network-Based Application Recognition (NBAR), traffic shaping (DTS), policing (DCAR), congestion avoidance (DWRED), weighted fair queuing (DWFQ), guaranteed minimum bandwidth (DCBWFQ), and so on

- **Multiservice** LLQ, Frame Relay Forum (FRF) 11/12, RTP header compression, Multilink Point-to-Point Protocol (MLP) with link fragmentation and interleaving (MLP/LFI)
- **Accounting** NetFlow export
- **Load Balancing** CEF load balancing
- **Caching** Web Cache Communications Protocol (WCCP)
- **Compression** Hardware and software compression
- **Multicast** Multicast distributed switching

You may not be familiar with all of these features, and we will not discuss all of them here. Instead, we'll explore some of the major things you should be aware of when considering running a feature on the RSP platform in distributed mode. We'll also look at some examples.

IOS Versions

Because of the VIP architecture, when a new feature comes out for the core router platforms, the code needs significant rewriting to port it to the RSP platform. As a result, a significant amount of time may pass before the feature is available in distributed mode, if at all, on the RSP platform. An example is RTP header compression (cRTP). Although this functionality was originally released on most platforms with IOS version 11.1(15), it was not released as a distributed feature on the 7500 Series until version 12.1(5)T. RTP header compression is a very processor-intensive service that would not scale to many connections on the RSP platform without distributed support.

Operational Differences

Although the underlying concepts of each of these features are basically the same in the distributed versions, the exact implementation may differ significantly. To maximize the efficiency of the router, the inner workings of a particular feature may need revision when running in distributed mode. Consider Cisco express forwarding as an example. With distributed CEF (DCEF), a forwarding information base (FIB) is built on the RSP, which caches the outgoing interface corresponding to a route. This FIB is shared by the RSP with each VIP by downloading it directly to the VIPs. Changes are made to the FIB only when the routing table changes. Any change to the FIB is again shared with all VIPs. Since the RSP and each VIP use the same FIB, there is an efficiency gain by sharing it.

This eliminates the per-flow overhead of route-cache maintenance. With DCEF, switching decisions are handled by the VIP whenever possible. However, if DCEF does not have the route cached, it will fall back to the next level of routing on the RSP, usually fast switching or process switching.

Restrictions

There may also be restrictions you need to be aware of when implementing features in distributed mode. There may be other features that need to be enabled before a particular feature is turned on, or there may be features that are disabled necessarily when a distributed feature is used. Distributed WRED (DWRED) is an example. WRED is used to avoid congestion by dropping packets and thereby throttling back congestion notification compliant flows. Distributed WRED was designed for ISP providers who would be aggregating many smaller circuits into larger ones. Since WRED needs to monitor the current buffer state of interface output queues, and because interfaces reside on the VIP, it makes sense that the WRED process would run in distributed mode on the VIP. Otherwise, the RSP would have to continually poll the status of all buffers across all VIPs. By offloading this function to the VIP, the main processor is freed up to do other things. However, at press time, there were some technical restrictions:

- You cannot configure DWRED on the same interface as RSP-based custom queuing, priority queuing, or weighted fair queuing (WFQ). You can, however, configure both DWRED and DWFQ (or DCBWFQ) on the same interface.
- DWRED is available only on a per interface basis. You cannot configure DWRED on a subinterface.
- DWRED is not supported with the ATM encapsulations AAL5-MUX and AAL5-NLPID.
- DWRED is not supported on Fast EtherChannel or Tunnel interfaces.

Although the goal is to make these distributed feature differences transparent and the restrictions as few as possible, differences and restrictions do exist. If you have 7500 Series routers in your network, it is up to you to do a little research to make sure the features you need can be extended to the RSP platform, and that they will scale by running them in distributed mode. There are many tools and a vast amount of documentation on Cisco's Web site that can help with this. The *IOS Feature Navigator* tool can help in determining if the feature you want is

available on a particular platform, and it can tell you what IOS code and feature set you need to run. Additionally, IOS software release notes can give you detailed information on how to configure distributed features. Finally, you can always contact your Cisco representative for especially difficult queries.

Using Link Fragmentation and Interleaving

Real-time and interactive traffic like Telnet, voice, and video can be affected negatively by jitter created when a router must process large packets on low-speed interfaces. These real-time streams usually consist of small packets, and jitter is caused when the regularly timed transmission of these packets is interrupted by the serialization delay of sending a large packet. Serialization delay is the fundamental time it takes a packet to be sent out a serial interface. It is based on the simple function:

$$\text{SerializationDelay} = \frac{\text{PacketSize(bits)}}{\text{BitRate(bps)}}$$

Table 8.5 shows serialization delays tabulated for common circuit speeds.

Table 8.5 Serialization Delays (Transmission Time for Link Speed [in ms])

Link Speed	64 Bytes	256 Bytes	512 Bytes	1024 Bytes	1500 Bytes
64	8	32	64	128	188
128	4	16	32	64	94
192	3	11	21	43	63
256	2	8	16	32	47
320	2	6	13	26	38
384	1	5	11	21	31
448	1	5	9	18	27
512	1	4	8	16	23
576	1	4	7	14	21
640	1	3	6	13	19
704	1	3	6	12	17
768	1	3	5	11	16

Continued

Table 8.5 Continued

Link Speed	64 Bytes	256 Bytes	512 Bytes	1024 Bytes	1500 Bytes
832	1	2	5	10	14
896	1	2	5	9	13
960	1	2	4	9	13
1024	1	2	4	8	12
1088	0	2	4	8	11
1152	0	2	4	7	10
1216	0	2	3	7	10
1280	0	2	3	6	9
1344	0	2	3	6	9
1408	0	1	3	6	9
1472	0	1	3	6	8
1536	0	1	3	5	8

Using a feature like LLQ or PQ can significantly reduce delays on real-time traffic, but even with this enabled, the time a real-time packet may have to wait for even one large packet to be transmitted could be large enough to add jitter to the stream. What usually happens is that after the priority queue empties, a large packet is started out the interface. Shortly after this, another packet comes into the priority queue but now has to wait for the whole large packet to be transmitted. Meanwhile, other priority packets queue up behind the first one at regular intervals. When the packets finally go, they go in a little burst. For an application like VoIP, the jitter buffer may have difficulty playing out all these packets smoothly with the delays and the bursts without dropping a packet or adding an unacceptably large amount of delay.

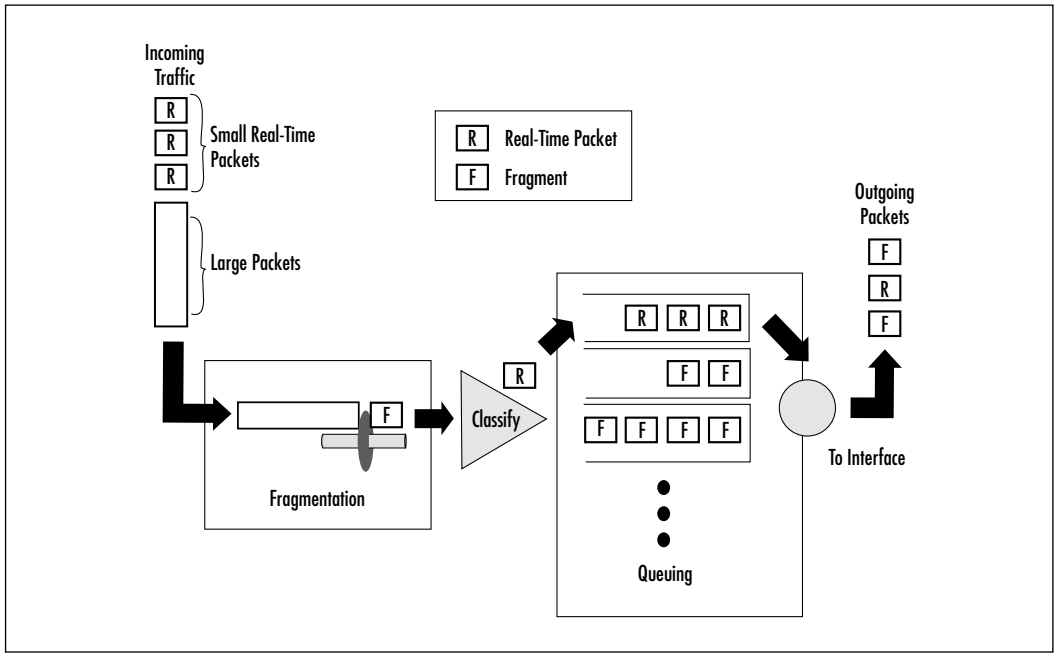
Link Fragmentation and Interleaving overcomes this by reducing the maximum packet size of all packets over a serial link to a size small enough that no single packet will significantly delay critical real-time data. These large packets that are broken up can now be interleaved with the real-time packets. LFI is superior to just changing the maximum transmission unit (MTU) size, because with LFI, fragmented packets are put back together at the other end of the serial

link. With MTU fragmentation, the packets travel across the whole network to their destination before being reassembled. This causes unnecessary traffic and processor utilization caused by increased header information.

How Does LFI Work?

Figure 8.6 shows the basic process of LFI. LFI effectively chops up large datagrams into smaller fragments (F) so they can be interleaved with real-time packets (R). These resulting smaller packets are then mixed in with other packets by whatever queuing mechanism has been configured (WFQ, LLQ, and so forth). When the packets arrive at the other end, they are reassembled into their original forms.

Figure 8.6 Link Fragmentation and Interleaving



How is the fragmentation size chosen? A particular packet size corresponds to a serialization delay. We choose the serialization delay by considering the maximum delay tolerated by the critical application. From this delay, the fragmentation size can be calculated as the product of the link speed and the target delay. Let us illustrate this with an example.

Imagine that we have a VoIP application running in the company of other data on a 128 Kbps circuit. Ethernet has an MTU of 1500 bytes, so on a 128

Kbps circuit without LFI, it would take 94 milliseconds (ms) to serialize the entire packet (refer back to Table 8.5). Therefore, a VoIP packet could potentially wait 94 ms before it could begin to be transmitted. This delay is too long and would cause jitter in the playout stream of the listener. VoIP is usually sent with two 10-ms samples in each packet. Assume we want to set the target delay between each packet to 10 ms. The fragmentation size for this circuit is thus calculated to be 160 bytes (128 Kbps x 10 ms). Therefore, to guarantee the target delay of 10 ms, each large packet needs to be fragmented into 160-byte pieces.

LFI with Multilink Point-to-Point Protocol

Multilink Point-to-Point Protocol (MLP) is an extension of PPP and is necessary for LFI to be used. It provides load-balancing functionality over multiple WAN links. With LFI, it provides packet fragmentation and proper resequencing, according to RFC 1717. It treats multiple links as one circuit and gives load calculation on both inbound and outbound traffic. Although MLP is necessary for LFI to be used, it is not necessary to have more than one WAN link.

NOTE

FRF.12 is the Frame Relay Forum standard for implementing LFI on Frame Relay. It was created with Voice over Frame Relay (VoFR) in mind, and like MLP with LFI, it must be used when voice packets are mixed with larger, nonvoice packets. It is available on the 2600/3600/MC3810/7200 platforms with IOS 12.0(4)T and later.

How Can This Be Useful on My Network?

If you are planning to implement on VoIP on circuit speeds less than 768 Kbps, LFI is indispensable, in an AVVID environment where your network is a true multiservice network and has data and video in addition to voice packets. At speeds below this, the MTU of an IP/Ethernet packet (1500 bytes) will take more than 15 ms to serialize. If you are really interested in good quality voice transmissions, you need to use some kind of priority queuing like LLQ as well. If your link speed is small, you will also obtain significant gains by another link efficiency mechanism, RTP header compression. We will consider that next.

Understanding RTP Header Compression

The general functions of RTP can be described as follows:

- Provides end-to-end network transport functions for audio and video over multicast or unicast services
- Supports real-time conferencing of groups
- Supports source identification of gateways and multicast-to-unicast translators
- Provides feedback from receivers on QoS
- Synchronizes video and audio streams with time stamping

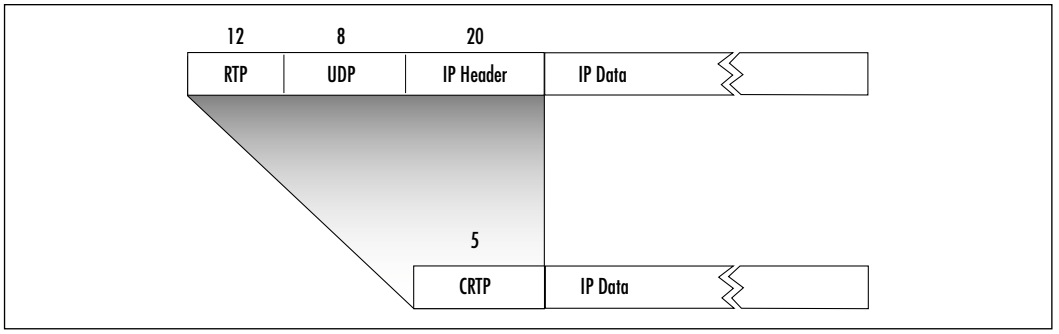
There has been a growing interest in RTP in AVVID environments because of its interoperability among different implementations of network audio and video applications. However, RTP has a header field of 12 bytes. This, combined with the encapsulated UDP and IP, increases the total overhead to 40 bytes. Because of the large amount of header information for the relatively small size of multimedia data payloads, extending RTP to slow links (dial-up modems, ISDN/BRI, subrate T1s) has been difficult. RTP Header Compression (CRTP) was created to offset the large header size associated with Real-Time Transport Protocol (RTP).

How Does RTP Header Compression Work?

RTP has two parts: a data portion and a header portion. The header can be much larger than the data portion and therefore can add quite a lot of overhead to media applications that use RTP. RTP is based on UDP, as opposed to TCP, since acknowledgements are not needed for real-time data streams. If a real-time data packet is lost, it does not usually make sense to resend it, since the time when it was needed would have already passed.

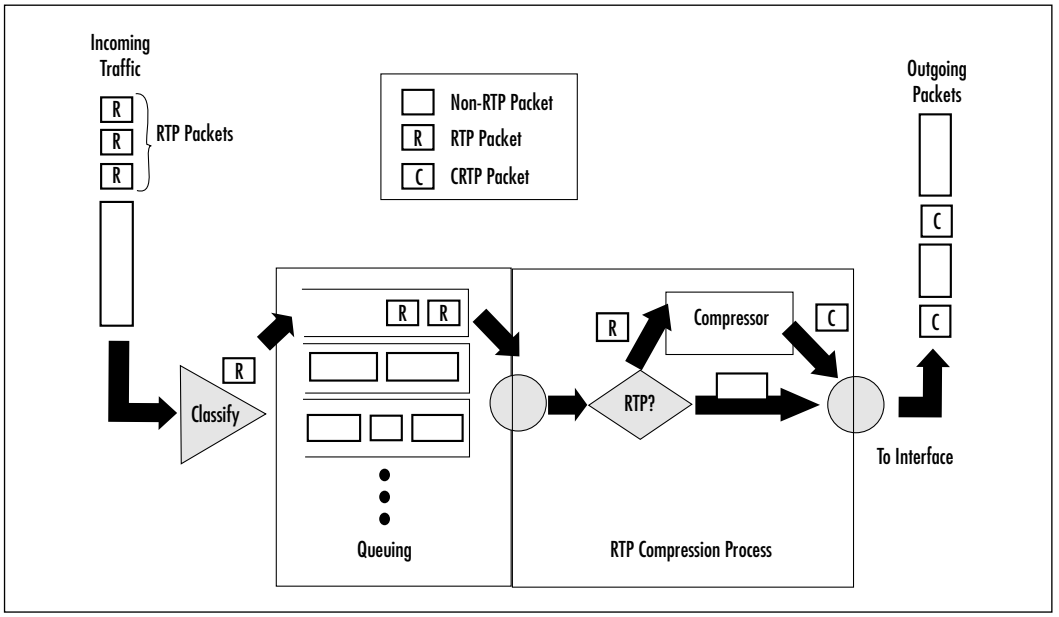
Since RTP encapsulates UDP and IP headers, the total amount of header information (RTP/UDP/IP) adds up to 40 bytes (see Figure 8.7). Considering the small packet size that usually comprises multimedia streams, this is a lot of overhead. Since most of the header information does not change very much from packet to packet, this lends to the idea of compressing it. RTP header compression can reduce this 40-byte header to about 5 bytes on a link-by-link basis.

Figure 8.7 RTP/UDP/IP Packet Headers



The RTP compression process is shown in Figure 8.8. We can see that CRTP works to compress RTP packets after the configured queuing process. Only RTP packets are compressed by the engine. After the compression process, both CRTP and non-RTP packets go to the interface to be serialized.

Figure 8.8 RTP Header Compression Process



When Would I Need RTP Header Compression?

RTP header compression can be useful on any narrowband link. Narrowband is usually defined by speeds less than T1, but in making your decision whether to use CRTP, you should consider not only your link speed, but also the available router resources (CPU) and your overall traffic patterns. Since CRTP is performed at the main processor, enabling it could cause your utilization to jump if you have high packet rates (lots of headers), many serial interfaces, or large serial interfaces.

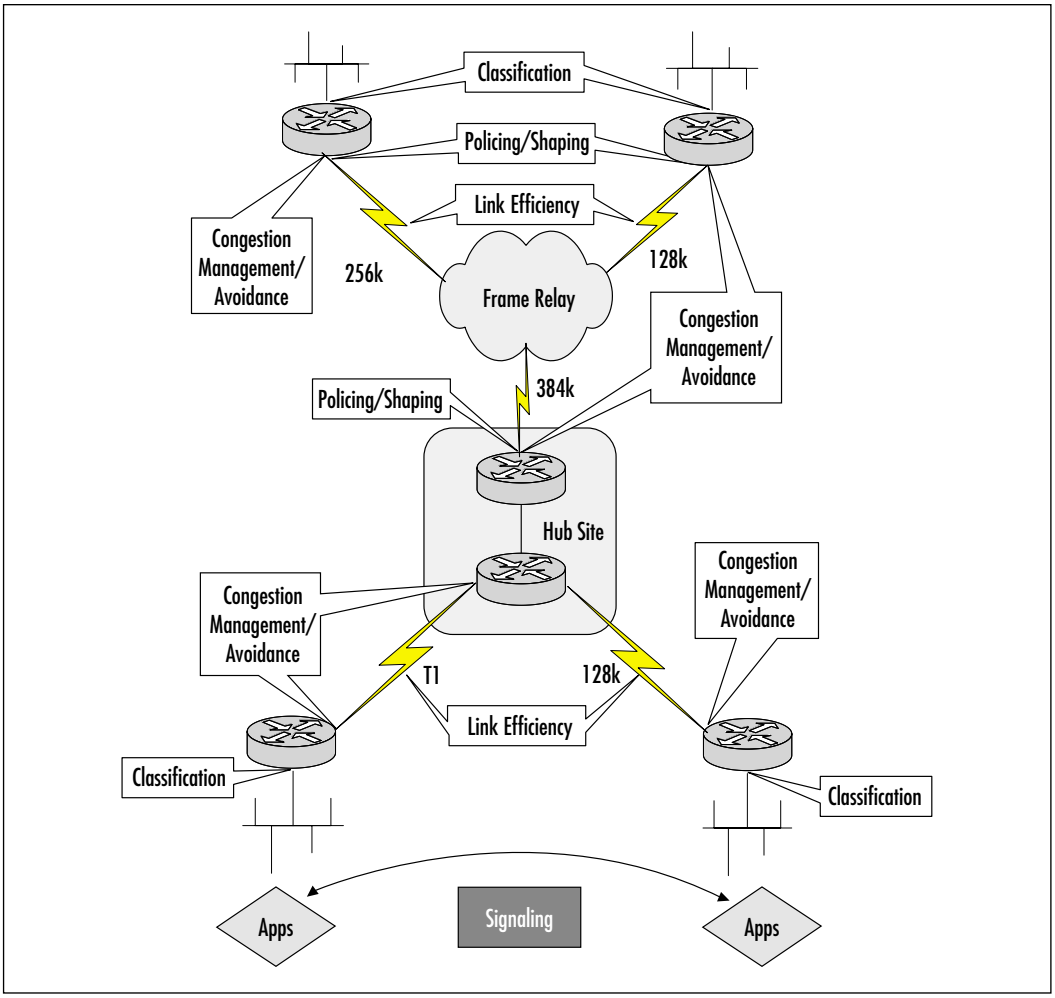
Summary

In this chapter, we introduced a lot of QoS mechanisms of various natures and explained how they can be used effectively to plan an overall QoS strategy for your AVVID environment. It is apparent that there is no single mechanism that is a cure-all for every situation. On the contrary, the power of these mechanisms lies in their capacity to work together. Each of the mechanisms has its own particular place in the network and its own particular function. *Classification*, simply put, can help to divide your traffic into classes of service, marking them so that other mechanisms can deliver differentiated QoS. With *Congestion Management* mechanisms, you can determine which packets get dropped and which ones get priority through the network. *Congestion Avoidance* works to prevent congestion by notifying senders to slow down when the network is busy. *Policing* and *Shaping* techniques regulate traffic flow according to set parameters, dropping traffic that does not conform, to avoid congestion situations downstream. *Signaling* can be used by clients to request end-to-end QoS across a network. Finally, *Link Efficiency* mechanisms can make the most efficient use of available bandwidth by using compression techniques and by binding small links into one logical pipe. The following list arranges the mechanisms discussed in this chapter by type and usage:

- **Classification/Marking** Used primarily at the edge of the network (ACLs, IP precedence)
- **Congestion Management** Used on serial interfaces (PQ, CQ, WFQ, CBWFQ, LLQ)
- **Congestion Avoidance** Used on Frame Relay interfaces and aggregating interfaces (RED, WRED, BECN, Foresight)
- **Policing and Shaping** Used on data-rate mismatched interfaces (GTS, FRTS, CAR)
- **Signaling** Used end-to-end between clients and on intermediate nodes (RSVP)
- **Link Efficiency** Used on low-speed and multilink interfaces (CRTP, MLP, LFI, FRF.12)

To help consolidate the information presented in this chapter, Figure 8.9 shows a small network indicating where each of these mechanisms would be applicable.

Figure 8.9 Advanced QoS Mechanisms in the Network



By classifying packets at the edge of the network into distinct classes, the network can provide differential services to the packets without having to examine each one in detail at every hop. After they are marked once with IP precedence, congestion management and avoidance mechanisms can act upon them as they travel to their destination. This is the essence of the DiffServ model of QoS. Since DiffServ does not employ end-to-end signaling between clients, it is basically a *connectionless* form of QoS. Although it may not be able to guarantee QoS totally, it will scale well and make efficient use of spare network resources.

On the other hand, we also looked at the Intserv model embodied by RSVP. Since RSVP is a signaling protocol between clients, it can be thought of as *connection-oriented* QoS model. Connection-oriented networks are traditionally good at providing QoS guarantees, but they do not make efficient use of spare bandwidth, and they have serious scaling problems.

So which approach should you take? There is no doubt that for most networks, DiffServ and IP precedence will be convenient and functional. Cisco's direction seems to indicate a growing trend towards creating and improving mechanisms that will heed class-marking bits such as IP precedence and Differentiated Services Code Point (DSCP). The lack of RSVP-enabled clients and other Intserv signaling mechanisms also points towards DiffServ. However, in future networks, the most effective QoS will probably be found by using a combination of the two models. One can imagine a large network (maybe even the Internet) with a DiffServ core, enhanced, perhaps, by technologies such as Multiprotocol Label Switching (MPLS), with an Intserv function like RSVP working at the client level. With technology changing as rapidly as it is, it is very difficult to predict.

Solutions Fast Track

Using the Resource Reservation Protocol

- ☑ RSVP does not provide QoS directly to applications, but instead, coordinates an overall service level by making reservation requests across the network. It is up to other QoS mechanisms to actually prevent and control congestion, provide efficient use of links, and classify and police traffic.
- ☑ End-to-end resource reservation can only be accomplished by using RSVP on every router end-to-end, but it is not mandatory that RSVP be enabled everywhere on a network. RSVP has the built-in capability to tunnel over non-RSVP aware nodes.
- ☑ Because of the resources required for each reservation, RSVP has some distinct scaling issues that make it doubtful it will ever be implemented successfully on a very large network, or on the Internet, in its current revision.

Using Class-Based Weighted Fair Queuing

- ☑ CBWFQ carries the WFQ algorithm further by allowing user-defined classes, which allow greater control over traffic queuing and bandwidth allocation.
- ☑ Flow-based WFQ automatically detects flows based on characteristics of the third and fourth layers of the OSI model. Conversations are singled out into flows by source and destination IP address, port number, and IP precedence.
- ☑ CBWFQ allows the creation of up to 64 individual classes plus a default class. The number and size of the classes are, of course, based on the bandwidth. By default, the maximum bandwidth that can be allocated to user-defined classes is 75 percent of the link speed.

Using Low Latency Queuing

- ☑ LLQ creates a strict priority queue that you can think of as resting on top of all other CBWFQ queues.
- ☑ LLQ overcomes the fact that low latency transmission may not be provided to packets in congestion situations, since all packets are transmitted fairly, based on their weight.
- ☑ Because of the nature of the LLQ, it is recommended that only voice traffic be placed in that queue.

Using Weighted Random Early Detection

- ☑ RED works on the basis of active queue management, and addresses the shortcomings of tail drop.
- ☑ WRED was primarily designed for use in IP networks dominated by TCP, because UDP traffic is not responsive to packet drop like TCP.
- ☑ WRED treats non-IP traffic as precedence 0, the lowest precedence. Therefore, non-IP traffic will be lumped into a single bucket and is more likely to be dropped than IP traffic. This may cause problems if most of your important traffic is something other than IP.

Using Generic Traffic Shaping and Frame Relay Traffic Shaping

- ☑ FRTS and GTS both use a token bucket, or credit manager, algorithm to service the main queuing mechanism and send packets out the interface. FRTS is commonly used to overcome data-rate mismatches.
- ☑ FRTS and GTS act to limit packet rates sent out an interface to a mean rate, while allowing for buffering of momentary bursts.
- ☑ Recall that queuing mechanisms will only kick in when there is congestion, so we need a mechanism to create congestion at the head-end. This is a common need on Frame Relay networks where the home office has much more bandwidth than any individual remote office.

Running in Distributed Mode

- ☑ When a process is run on the VIP instead of the main processor, the service is said to be running in *distributed mode.*
- ☑ Most of the QoS features you will find useful in an AVVID environment were introduced (in distributed mode) in 12.1(5)T.

Using Link Fragmentation and Interleaving

- ☑ Real-time streams usually consist of small packets, and jitter is caused when the regularly timed transmission of these packets is interrupted by the serialization delay of sending a large packet. Serialization delay is the fundamental time it takes a packet to be sent out a serial interface.
- ☑ Using a feature like LLQ or PQ can significantly reduce delays on real-time traffic, but even with this enabled, the time a real-time packet may have to wait for even one large packet to be transmitted could be large enough to add jitter to the stream.
- ☑ Link Fragmentation and Interleaving overcomes this by reducing the maximum packet size of all packets over a serial link to a size small enough that no single packet will significantly delay critical real-time data.

Understanding RTP Header Compression

- ☑ RTP encapsulates UDP and IP headers, and the total amount of header information (RTP/UDP/IP) adds up to 40 bytes. Since small packets are characteristic of multimedia streams, that is a lot of overhead. Most of the header information does not change from packet to packet, so RTP header compression can reduce this 40-byte header to about 5 bytes on a link-by-link basis.
- ☑ RTP header compression can be useful on any narrowband link. Narrowband is usually defined by speeds less than T1.
- ☑ Since cRTP is performed by the main processor, enabling it could cause your CPU utilization to jump if you have high packet rates, lots of serial interfaces, or large serial interfaces. Use this feature with caution.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: Is CBWFQ better than PQ, CQ, or basic WFQ?

A: WFQ is a great queuing mechanism when you want every packet flow to have fair use of the available link bandwidth. PQ can give better than fair treatment for particular classes, but at the risk of starving some traffic of bandwidth completely. CQ allows guarantees of bandwidth to particular types of traffic, but it may be too granular, since unclassified flows are all given the same treatment. CBWFQ works on top of WFQ to provide flow-based prioritization with bandwidth reservations for the classes you specify, and is probably the best choice in an AVVID environment because CBWFQ/LLQ includes a strict priority queue for real-time packets (the Low Latency Queue), which is specifically intended for voice packets.

Q: How do I use RSVP with non-RSVP-enabled clients?

A: For the most part, you cannot. Because RSVP is end-to-end signaling between clients, the clients themselves must be able to send RSVP packets. However, if you are using Cisco routers as voice gateways, you can set up VoIP dial-peers to use RSVP in order to request *controlled-load* or *guaranteed-rate* service. It is also possible to set up a reservation for a non-RSVP client using the RSVP proxy function on Cisco routers, but this is mostly for testing purposes, since each reservation must be manually configured, and the reservation will stay in place until the configuration is removed.

Q: While testing a Cisco AVVID solution, I have enabled LLQ in my lab for VoIP traffic. To a simulated remote site, I have reserved enough bandwidth to support 12 simultaneous calls. When more than 13 calls are connected, call quality becomes poor. How do I resolve this?

A: The short answer to this is "Admission Control." When the strict priority queue that is behind low latency queuing reaches its configured bandwidth for priority packets, if congestion exists on the line, packets in excess of this configured bandwidth will be dropped on the 7200 Series and lower-end platforms. Because packets may be dropped from all voice streams, the quality of all calls can be affected. Admission control acts to limit the number of calls from one area to another, or across a link, to a particular number. In this way, the thirteenth simultaneous call would be refused by a busy signal or rerouted to the PSTN. Admission Control can be implemented in a number of ways including using Cisco Call Manager and gatekeeper software on Cisco routers.

Q: When do I need LFI?

A: LFI is needed only on serial links below a speed of 768 Kbps when small, real-time packets such as voice are being transmitted with other large packets such as FTP.

Q: When should I use RTP header compression?

A: In general, RTP header compression should be used for links below T1/E1 speed where there is a reasonable number of RTP packets on the link. The more RTP traffic such as voice and video you have on the link, the more benefit you will get from CRTP. CRTP is processor-intensive, so you will want to limit the number of RTP compressed links on a single router to a value appropriate for that router's processor speed.

Q: What are the best QoS mechanisms to use in my AVVID environment?

A: There is no simple formula for determining the best mechanisms for a particular network. You will likely want to become familiar with the features of each mechanism and use a combination of some or all of them, depending on your network and your business requirements. These mechanisms each have their own place in the network and work together to provide QoS.

Chapter 9

AVVID Dial Plans

Solutions in this chapter:

- **What Is a Dial Plan?**
- **Cisco CallManager Dial Plans**
- **Creation of Calling Restrictions and Configuration of Dial Plan Groups**
- **Guidelines for the Design and Implementation of Dial Plans**
- **The Role and Configuration of a Cisco CallManager and Gatekeeper**
- **Video Dial Plan Architecture**

- ☑ **Summary**
- ☑ **Solutions Fast Track**
- ☑ **Frequently Asked Questions**

Introduction

As with other parts of designing an AVVID network using Cisco CallManager, a *dial plan* should be a main consideration during the initial phase of the network design. You should make sure you have the correct dial plan that's suitably designed the first time to negate the need to change it once it's fully implemented. These changes can be very time- and resource-consuming and may lead you into areas where you will not be able to receive any support.

This chapter should give you an understanding of what a dial plan is and why it is so important to the creation and support of your network infrastructure. It will also explain the procedures of dial plan design for voice and video networks.

Problems Facing the Integration of Voice and Data

A major problem that has slowed integration deployment of the telephony and data networks is that there currently isn't a standardized dynamic routing protocol able to publicize dial plans between multiple telephony devices the way Open Shortest Path First (OSPF) and Enhanced Interior Gateway Routing Protocol (EIGRP) can in Cisco routers. In other words, these protocols are not currently designed to handle the transport of this information, meaning other protocols must first arise more specifically designed for voice and data architectures. Within the traditional telecommunications infrastructures, switches and routers are manually updated (there is limited automation) to reflect the routes located within the device. Though this has led to a very stable platform, it also lends itself to administration nightmares, as well as supreme support issues.

With voice and video networks, each destination and forwarding device along a path needs to have a dial plan. In the data and packet networks, this consists of implementing static routes, with the destination prefixes or even the direct destination added to the routing table. As you probably know, static routes offer minimal overhead on the network equipment since they basically act as a conduit for information, but there is usually massive support overhead for problems that arise. These problems are most often associated with the configuration (or misconfiguration) of routers. This can affect the entire network if even one of the routers along the path is incorrect. Should this occur, calls will not be able to terminate at the proper destination, if at all.

There is a strong push to implement a dynamic routing protocol for telephony devices, such as Simple Gateway Control Protocol (SGCP), Media

Gateway Control Protocol (MGCP), Session Initiation Protocol (SIP), and the newest, Telephony Border Gateway Protocol (TBGP).

Until these protocols gain acceptance and standardization, you will find that all dial plans on Cisco voice-enabled routers need to be implemented through dial peers. Dial peers are those commands used to configure the ports on routers with an associated phone number. This will help the router to determine what path is needed to complete the call.

What Is a Dial Plan?

So what is a dial plan? A dial plan is, in its most basic form, a system interface for telephony devices that allows users and equipment to connect to each other by using dialing strings. These dial strings can be mapped and routed to a multitude of locations by the controlling system. Think about it like this: a dial plan is an internal address you can route to in order to get from point A to point B. Without a dial plan, your telephony device would not be able to connect to another (granted, this might happen anyway, but you want to give yourself at least a shot). A dial plan is an integral part of the puzzle and gives your users the flexibility and services not normally associated with a standard PBX.

Designing & Planning…

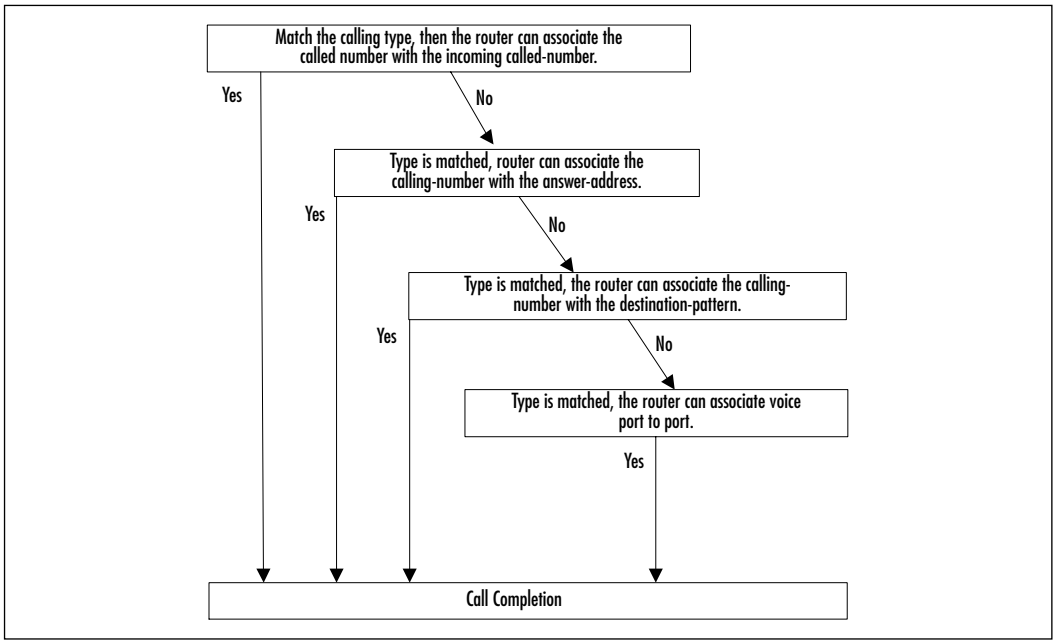
Simplicity Is the Key

One of the things to keep in mind when deploying a dial plan is simplicity. You want the setup to be as simple as possible to minimize the number of entries input into the Cisco CallManager (discussed later) and the router. By configuring the system in this manner, you are able to add a higher level of efficiency to the system, which can offer you better performance and less support headaches.

To this end, there are some things that are just common sense, such as only configuring dial plans for locations on the network (On-Net), and calls that are destined to go externally (Off-Net) to use the PSTN gateway access code (usually a 9, but can be defined). You will also want to create contingency plans for internal calls to complete through the PSTN if the network is congested. One thing that customers will always demand is dial tone.

Routers, when used in an AVVID implementation, are like filters. They work to match dial peers. When a call is arriving on a POTS port, the router will match the VoIP dial peer for the outbound call. Conversely, this is also true when calls come to the router from an IP network. The router will match both the POTS peer (for call termination) and the VoIP peer to utilize AVVID bandwidth features like Voice Activity Detection (VAD). This can be easily translated into flow chart form like that shown in Figure 9.1.

Figure 9.1 Matching Dial Peers



Using this same logic flow, the router could also match incoming calls from the POTS to POTS dial peers. When configuring dial plans, you will need to manually define them on the Cisco Routers using dialing peers. Dialing peers are like static routes, helping the router make decisions on where calls can originate and terminate, and what paths can be used. The attributes you configure will help determine how the call is placed when the user dials their phone.

When used in conjunction with AVVID, a dial plan allows you to create your network so it accommodates data and voice/video within your infrastructure. A dial plan can be created in many different ways depending on the needs that arise for each individual deployment. There can be plans that include redundant paths, such as IP networks and the PSTN so any calls made have a path to completion.

You can incorporate abbreviated dialing, allowing offices to use minimal dialing to complete intraoffice calls or even reroute calls. Or you can create call restrictions, such as blocking 900 numbers.

NOTE

If you are implementing MGCP or SGCP as your calling agent, you do not need to statically configure dial peers. If you are in an environment where you are configuring multiple routers for your calling cluster, you will have to configure the dial peers on all associated routers.

Configuring Dial Peers for Use

Configuring dial peers is essential when designing and implementing Voice over IP on your network. Dial peers identify the calling source and the destination points so as to define what attributes are assigned to each call. In the telecommunications world, calls routed over the PSTN are assigned to a dedicated 64KB circuit from start to end. In the data world, a voice call must traverse segments within the network referred to as call legs. This isn't to say that these don't exist in the telecomm world, they are just more noticeable within packet-based networks.

A call leg is the connection that occurs between the calling device and the router, router-to-router along the path, and from the router to the end device. A dial peer is linked with each of these segments, and it's here where the defined attributes are added to the call. Things like the CODEC, VAD, and QoS are utilized depending on whether you have defined them for that link.

There are several types of dial peers, but those we'll focus on here are VoIP and POTS. VoIP dial peers are associated with the IP address of the destinations router so it can connect and terminate a call from an IP-based telephony device. POTS dial peers are basically the telephone system as we know it. These dial peers map dialed digit strings to specific voice ports that are located on the router. These voice ports are usually associated with the PSTN and the PBX.

Configuring Dial Peers for POTS

When you configure a dial peer for POTS, you need to assign a unique tag to the dial peer for identification, define the destination (either a telephone number

or a range of telephone numbers), and then associate a voice port so calls can be established. Table 9.1 shows the standard configurations.

Table 9.1 Basic Configuration Commands for the Creation of Dial Peers

Command	Description
Gatekeeper(config-dial-peer)# voice *number* **pots**	This allows you to enter dial-peer configuration mode so you can define a local dial peer connecting to a POTS interface. The argument *number* is one or more digits that can identify the dial peer. Valid entries are from 1 to 2147483647. The keyword *pots* specifies that the dial peer is using basic telephone service.
Gatekeeper(config-dial-peer)# **destination-pattern** *string* [T]	This command matches dialed digits to a telephony device. The argument *string* is a series of digits that specify the addressing is E.164, or a private dialing plan telephone number. Valid entries are the numbers 0 through 9 and the letters A through D. When the timer (T) character is included at the end of the destination-pattern, the router then collects dialed digits until the interdigit timer expires, approximately ten seconds if left at the default, or until you dial the termination character. Usually this is left as the default. Be aware the timer character must be a capital T for it to work.
Gatekeeper(config-dial-peer)# **port** location	This command maps a dial peer to a specific logical interface that it needs to be associated with. Be aware that the **port** command syntax is platform-specific.

Options for the Configuration of Dial Plans for POTS Dial Peers

There are also some configurable options to help you shape the deployment of your dial peers. Table 9.2 is a list of some of the most common customization commands.

Table 9.2 Optional Dial Plan Configuration Commands

Command	Description
Gatekeeper(config-dial-peer)# **answer-address** *string*	(Optional) This command selects the inbound dial peer based on the calling-number.
Gatekeeper(config-dial-peer)# **incoming called-number** *string*	(Optional) This command selects the inbound dial peer based on the called-number so it can identify voice and modem calls.
Gatekeeper(config-dial-peer)# **direct-inward-dial** *string*	(Optional) This command enables Direct Inward Dialing (DID) call treatment for the incoming called-numbers.
Gatekeeper(config-dial-peer)# **forward-digits** { *num-digit* | **all** | **extra**}	(Optional) This command configures digit-forwarding for the dial peer. The valid range for the number of digits that can be forwarded *(num-digi*t) is 0 through 32.
Gatekeeper(config-dial-peer)# **max-conn** *number*	(Optional) This command specifies the maximum number of connections allowed to and from the POTS dial peer. The valid range is 1 through 2147483647.
Gatekeeper(config-dial-peer)# **numbering-type {abbreviated | international | national | network | reserved | subscriber | unknown}**	(Optional) This command specifies which numbering type to match, as defined by the ITU Q.931 specification.
Gatekeeper(config-dial-peer)# **preference** *value*	(Optional) This command configures the preference for the POTS dial peer. The valid range is 0 through 10. The lower the number, the higher the preference is for that dial peer.
Gatekeeper(config-dial-peer)# **prefix** *string*	(Optional) This command adds a prefix that the system will prepend to the dial string before passing it to the telephony interfaces. Valid entries for the *string* argument are 0 through 9 and a comma. You would use the comma to include a one-second pause between digits to allow for a secondary dial tone.

Continued

Table 9.2 Continued

Command	Description
Gatekeeper(config-dial-peer)# **translate-outgoing {called** | calling} *name-tag*	(Optional) This command specifies the translation rule set that is applied to the calling-number or called-number.

Configuring Dial Peers for VoIP

By configuring VoIP dial peers, you can enable the router to make outbound calls to other telephony devices, located within the network. In order to configure a dial peer for VoIP, you need to assign a unique tag to the dial peer for identification, define the destination telephone number, and define the destination IP address. Table 9.3 shows some basics.

Table 9.3 Dial Peer Commands for Implementing VoIP

Command	Description
Gatekeeper(config-dial-peer)# **dial-peer** voice *number* **voip**	This command enters dial-peer configuration mode and will define a remote VoIP dial peer. The argument *number* is one or more digits used to identify the dial peer. Valid entries are from 1 to 2147483647. The keyword *voip* specifies a dial peer that uses voice encapsulation on the IP network.
Gatekeeper(config-dial-peer)# **destination-pattern** *string* [T]	This command configures the dial peer's destination-pattern so the system can resolve dialed digits with a telephone number. The argument *string* is a series of digits used to specify the addressing is E.164, or the private dialing plan telephone number. Valid entries are the numbers 0 through 9 and the letters A through D.
Gatekeeper(config-dial-peer)# **session target {ipv4: destination-address | dns:[s. | d. | e.| u.]** *host-name*}	This command is used to define the IP address of the router, which is connected to the remote telephony device.

Continued

Table 9.3 Continued

Command	Description
	The keyword and argument *ipv4: destination-address* indicate the IP address of the remote router. The keyword and argument *dns:host-name* indicates that the domain name server will resolve the name of the IP address. Valid entries for this parameter are characters representing the name of the host device. Wildcards are also available for defining domain names with the keyword by using source, destination, and dialed information in the host name.
Gatekeeper(config-dial-peer)# **codec {g711alaw | g711ulaw | g723ar53 | g723ar63 | g723r53 | g723r63 | g726r16 | g726r24 | g726r32 | g728 | g729br8 | g729r8 [pre-ietf]}** [*bytes*]	This command defines the CODEC for the dial peer. The optional switch *bytes* will set the number of voice data bytes per frame. Values are from 10 to 240 in increments of 10 (for example, 10, 20, 30, and so on) are considered acceptable. Any other value is rounded down (for example, from 144 to 140). The CODEC value must be matched on both VoIP dial peers on either side of the connection. If you specify g729r8, then IETF bit-ordering will be used. Be aware that the CODEC command syntax is platform- and release-specific.

Options for the Configuration of Dial Plans for VoIP Dial Peers

There are also some configurable options to help you shape the deployment of your dial peers. Table 9.4 is a list of some of the most common customization commands.

Table 9.4 Optional Commands for the Configuration of VoIP

Command	Description
Gatekeeper(config-dial-peer)# **answer-address** *string*	(Optional) This command chooses the inbound dial peer based on the calling-number.
Gatekeeper(config-dial-peer)# **incoming called-number** *string*	(Optional) This command chooses the inbound dial peer based on the called-number, to identify voice and modem calls.
Gatekeeper(config-dial-peer)# **dtmf-relay** [cisco-rtp] [h245-signal] [h245-alphanumeric]	(Optional) This command is used to configure the tone that sounds in response to a pressed digit on a touch-tone telephone. Dual Tone Multi-Frequency (DTMF) tones are compressed at one end of a call and decompressed at the other. Be aware that if you use a low-bandwidth CODEC, such as G.729 or G.723, the tones can sound distorted, which may lead to problems. The **dtmf-relay** command transports DTMF tones generated after call establishment out-of-band. It uses a method that sends with greater reliability than what is possible in-band for most low-bandwidth CODECs. Without DTMF Relay, calls established with low-bandwidth CODECs may experience trouble accessing automated telephone menu systems such as voice mail and Interactive Voice Response (IVR) systems. A signaling method is supplied only if the remote end supports it. Options are the Cisco proprietary Real Time Protocol **(cisco-rtp**), standard H.323 **(h245-alphanumeric**), and H.323 standard with signal duration **(h245-signal**).
Gatekeeper(config-dial-peer)# **fax rate {2400 | 4800 | 7200 | 9600 | 12000 | 14400 | disable | voice}**	(Optional) This command indicated the transmission speed of a fax to be sent to this dial peer. The keyword **disable** turns off fax transmission capability. The keyword **voice**, which is on by default, specifies the highest possible transmission speed supported by the voice rate.

Continued

Table 9.4 Continued

Command	Description
Gatekeeper(config-dial-peer)# **numbering-type {abbreviated | international | national | network | reserved | subscriber | unknown}**	(Optional) This command indicates the numbering type to match, as defined by the ITU Q.931 specification.
Gatekeeper(config-dial-peer)# **playout-delay mode {adaptive | fixed}**	(Optional) This command indicates the type of jitter buffer playout delay to use.
Gatekeeper(config-dial-peer)# **playout-delay {maximum value | nominal** *value* **| minimum {default | low | high}}**	(Optional) This command indicates the amount of time a packet will be held in the jitter buffer before it is played out on the audio path.
Gatekeeper(config-dial-peer)# **preference** *value*	(Optional) This command configures the preference for the VoIP dial peer. The value is a number from 0 through 10. The lower the number, the higher the preference.
Gatekeeper(config-dial-peer)# **tech-prefix** *number*	(Optional) This command indicates a particular technology prefix that will be prepended to the destination-pattern of this dial peer.
Gatekeeper(config-dial-peer)# **translate-outgoing {called | calling}** *name-tag*	(Optional) This command indicates the translation rule set that needs to be applied to the calling-number or called-number.
Gatekeeper(config-dial-peer)# **vad**	(Optional) This command enables voice activity detection (VAD). This will disable the transmission of packets during periods of silence. VAD is on by default. The minimum time of silence detection for VAD can be configured by using the **voice vad-time** global configuration command. The music threshold can be configured by using the **music-threshold** voice-port command, if you feel it is affecting VAD performance.

Dial Peers for Inbound and Outbound Calls

Inbound and outbound calls use dial peers to receive and complete calls. You must remember that the definition of inbound and outbound is based on the perspective of the router. What this means is that a call coming into the router is considered an inbound call while a call originating from the router is considered an outbound call.

When an inbound call is destined for a device on the packet network and is coming from a POTS interface, the router will match the dial peers for the voice network with the inbound call leg so it is properly routed to the outbound port. If the call originates within the packet network, then the router will match the POTS dial peer and a voice network dial peer so it can modify its attributes for VAD, CODEC, and QoS.

Routers that receive inbound POTS calls are destined for outbound voice network dial peers, it will forward all of the collected digits. For outbound POTS calls, the router will remove explicitly matched digits and forward the remaining digits to the destination port.

The following configuration is a basic example of POTS and VoIP peers:

```
dial-peer voice 1 pots
destination-pattern 707....
port 1/0:1
dial-peer voice 2 voip
destination-pattern 707....
session target ipv4:10.1.100.1
```

As you can see, the router will choose a dial peer for a call leg by matching the digits defined by the destination-pattern, but it can also use the answer-address or incoming called-number commands if they are used within the dial peer configuration. Be aware that the character "." is the only wildcard applied if you use answer-address or incoming call-number commands for the creation of your dial peers.

Usage of the Destination-Pattern

To associate a dialed string with a specific telephony device, you would use the destination-pattern. With it, the dialed string will compare itself to the pattern and then be routed to the voice port or the session target (discussed later) voice network dial peer. If the call is an outbound call, the destination-pattern could also be used to filter the digits that will be forwarded by the router to the

telephony device or the PSTN. A destination-pattern must be configured for each and every POTS and VoIP dial peer configured on the router.

You could describe the destination as an entire number or just a partial number with digits that can be defined through the wildcard switch. The wildcard digit "." represents an individual digit the router will be expecting to receive. If a destination patter is defined as 707…., then all dialed digits that start with 707 and have four following digits will match this dial peer.

The "." is not the only character that can be used to represent other digits. Several others are listed in Table 9.5, along with a brief description, to assist you in the configuration of your dial peers.

Table 9.5 Character Representations

Character	Description
.	This character represents a single digit. Ex 707…. (where …. equals four following digits).
[]	These characters represent a range of digits. If the – is used such as [4–7] then the digits will be consecutive. If a comma is used, like in [4,7], then the range is nonconsecutive. You can also use a combination of each [4–7,9]. Note: this only works for single digits [4–7] not [37–41].
()	These characters represent a pattern, 425(707). They are normally used with the ?, %, and/or the +.
?	This character is used to specify that the previous digit happened zero or one time(s) (to use this character you must use the Ctrl+v key combination).
%	This character is used to specify that the previous digit happened zero or one time(s). It acts like an asterisk (*) and is used in a regular expression.
+	This character specifies that the previous digit occurred one or more times.
T	This character specifies the timeout used by the interdigit command.
* or #	These characters are standard on touch-tone telephones and can be used within the dial pattern or as a signal that the user is done dialing digits using the dial-peer terminator command.
$	This character, when used at the end of a dial string, will disable variable-length matching for the dial pattern.

The Session Target

The session target is the IP address of the router to which the call will be directed once the dial peer is matched. In a VoIP network, you need to configure this using the session target command under the destination-pattern configuration. For dial peers that are outbound, the destination-pattern is the telephone number associated with the device you want to connect to. On inbound dial peers, the session target is ignored.

Route Pattern (On-Net)

If you are working with multiple sites across a Wide Area Network with connections like frame or dedicated circuits, you have the ability to implement on-net Calls. On-net calls are when you make a call that remains within the network infrastructure. When using on-net, you have the ability to use abbreviated dialing string in order to complete calls to other offices. This is just for ease of dialing to the end user. As an example, let's say you have an office in Seattle that has a number range of (206) 707-0000 through (206) 707-0999. You would only need a single route pattern of 70XXX to complete a call to the Seattle office. The benefit of this is that it only requires one route pattern entry since the Xs work as wildcards.

The Cisco CallManager will use route patterns to add or remove digits to the dialed number. The reason for this is that all dialed strings filtered though the CallManager must have the appropriate number of digits in order to reach remote sites (even those located on the same WAN). The Cisco CallManager simply routes the calls based on these addresses. This is also done to make sure incoming call numbers don't need to be changed.

If the WAN cannot complete calls (either due to no connectivity or lack of sufficient bandwidth), the call will be routed over the PSTN (yet another reason for the route patterns). In some instances, you will need to have an area code added to the dial-string. When Cisco CallManager was first released, it was only able to prepend one set of numbers to any dialed string. Because of this, you had to use the Cisco IOS gateway to insert the area code (and in some instances, the three-digit exchange). Cisco fixed that with the release of Cisco CallManager 3.0, which can now add or remove numbers based on a per-route-group basis. So, you can now manage the entire system from one centralized point that can control the Cisco IOS gateways (and gateways that use the Skinny Gateway protocol as well).

Routing Outbound Calls through the PSTN

Calls destined to be routed through the PSTN usually require only one route pattern. In some offices, you may find it necessary to create an access code to access the PSTN, such as dialing a 9 before the number. In North America, the dialing convention is divided into sections. There is an area code (510), the exchange number (536), and the station ID number (XXXX). In order to make a long distance call (a call outside your calling area), you may also need to dial a 1 at the beginning of the string. In some cities the convention for ten-digit dialing is always necessary to complete calls. In these circumstances it is necessary to dial the area code, but not the preceding 1.

With Cisco CallManager, you are able to create route patterns allowing you to route calls that differentiate between a local call that requires ten-digit dialing and a call that only requires seven-digit dialing. If the rule is not set, then Cisco CallManager will wait ten seconds without dialed digit detection, and will assume if there are no other digits dialed, then the user has completed dialing.

Creation of a local PSTN gateway dial plan is easy (and mostly painless). Gateways that are based on Skinny Gateway Protocol and MGCP will have their dial plan information configured within Cisco CallManager itself, whereas H.323 gateways will require only a small set of dial peers. The dial peers are then used by the gateway to direct calls destined for the PSTN through the Cisco CallManager.

If you are located outside North America, the numbers of digits that must be dialed for call completion differ. In this case, you will need to create multiple length dial-plans. The problem is, with the current version of Cisco CallManager, the system doesn't know when the dialing is complete, so you need to create specific route patterns.

Cisco CallManager Dial Plans

By using Cisco CallManager, you are able to allow for greater growth and functionality within your network because it was designed to be integrated with Cisco's Internet Operating System (IOS) gateways.

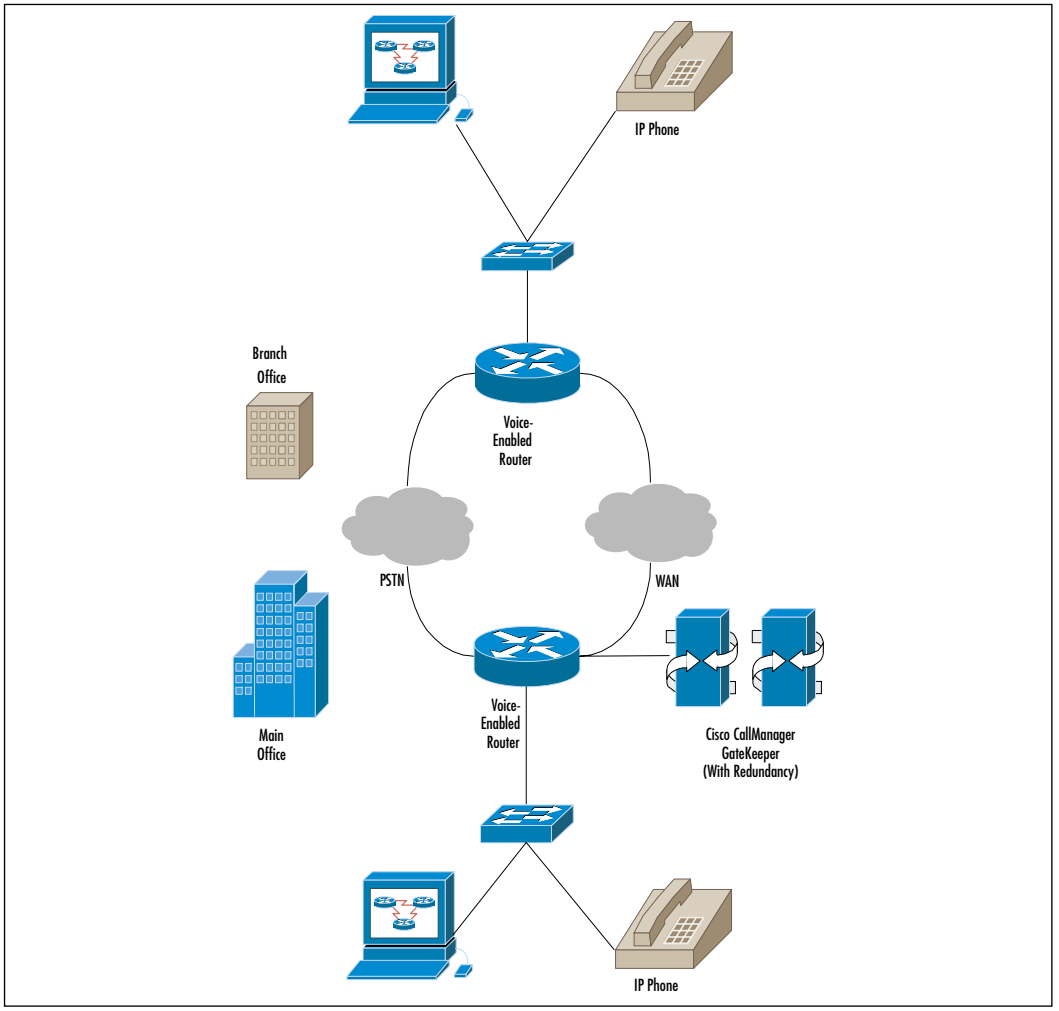
Cisco CallManager dial plans are usually created to handle two types of calls, internal and external:

- Internal calls are those calls initiated and terminated on Cisco IP phones that are included (registered) to the Cisco CallManager cluster.

- External calls are those calls passed through a PSTN gateway or a Cisco CallManager that originate across a WAN connection.

Figure 9.2 is a network designed to handle calls destined for the WAN and the PSTN. For this setup, voice calls would set the preference for the WAN and would only be routed to the PSTN if the WAN were down or unavailable. This routing takes place transparently to the user. In Figure 9.2, the Cisco CallManager Gatekeeper is a router assigned to manage this specific task as a gatekeeper. This router could also handle other items, but often it is best to have the router taking care of just Gatekeeper functions.

Figure 9.2 Simplicity and Redundancy



Designing & Planning…

Dial Plan Preferences

It is generally considered a good idea to create a dial plan that preferences certain paths routed across the IP network. If this network becomes unavailable, then calls should be routed across the PSTN. As always, the process should be transparent to the user.

Internal Calls

The creation of dial plans for internal calls to IP phones registered within a Cisco CallManager cluster is very simple. When the phone is initially configured, it is assigned a directory number (DN). This DN is maintained throughout the configured life of the phone. For example, if the phone is used in an office where your users move frequently within the LAN, their phones can be unplugged and connected to a different network jack, yet maintain their connection properties (DN). When the phone is reconnected, it will update the Cisco CallManager with its new IP address.

Designing & Planning…

The Mobility of IP Devices

IP phones are not the only network devices that work with DN connection properties. Cisco CallManager will also maintain the DN with Cisco IP SoftPhones, and certain types of analog devices (such as phones and facsimile machines) connected to gateways that use MGCP or the Skinny Gateway Protocol.

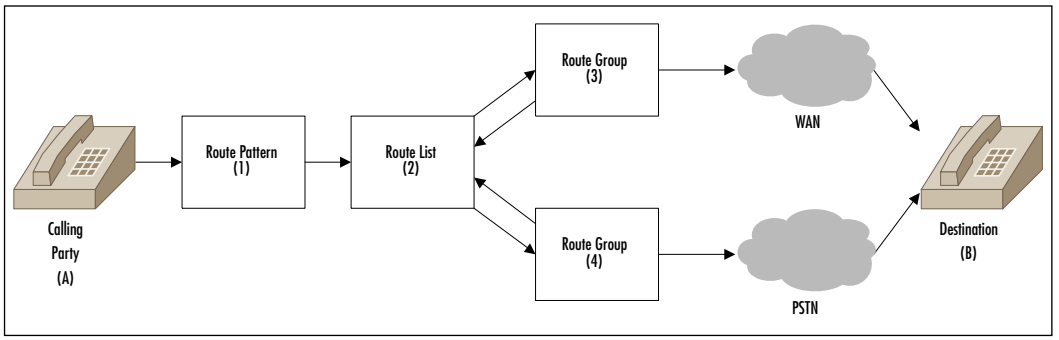
External Calls

Configuring Cisco CallManager to complete external calls requires implementing a *route pattern*. A route pattern is used to direct calls off network to a PSTN gateway. Route patterns can also be used if there are Cisco CallManagers located on a WAN-connected network.

Cisco CallManager dial plans are usually deployed in a tier system. This system lets different routes handle dialed numbers. You can also manipulate dialed strings, based on network requirements. This manipulation can either add or subtract digits from the number dialed by the user so as to accommodate network and gateway needs. Cisco CallManager can also create Trunk groups that will handle redundancy and create better paths for least-cost routing. For example, when using trunk groups, the system has the ability to choose an alternate route to complete (or in some cases admit) calls if the trunks do not have sufficient bandwidth to handle the call. This can be denoted (when creating the dial plan) as a continuation of the rule that moves calls to the PSTN if WAN connections are saturated.

In Figure 9.3, a call is placed from a telephony device (A). It is then matched against the route pattern (1), where digit manipulation takes place. From here, the call is forwarded to the route list (2). The route list adds the preference of connecting the call over the WAN link. If the call is unable to be completed through the WAN (because of insufficient resources or some other reason), then the call will be forwarded to the PSTN. If the PSTN cannot complete the connection, then the user will receive a busy signal (unless there is a third route configured). From either the WAN or the PSTN, the call is forwarded to the destination party (B). Again, this entire process should be transparent to the end user.

Figure 9.3 Flow of a Call through a Cisco CallManager Route Pattern



Route Pattern

A route pattern is the addressing method that identifies the dialed number and uses route lists and route group configurations to determine the route for call completion. Dialed numbers (E.164 North American Standard) are broken down into smaller groups, creating route patterns that can be entered into the Cisco CallManager as a specific number (for point-to-point direct dialing) or as a number range (the more common implementation). By using a route pattern, you can summarize a large range of numbers so minimal entries are needed to route a call.

As a dialed number is routed, the CallManager will look to create a pattern match, so the call can be correctly routed to the next hop and eventually to the end devices. Keep in mind, the digits can still be changed by the CallManager before they are put into the route list. By this method, numbers can be added or subtracted to the dialed strings. Once the number is passed to the route list, it will determine which route it will take to its next route groups (also trunk groups) and prioritize the traffic and connections.

What Is Digit Manipulation, and How Do You Configure It?

In the real world, you may have a device that is already in use, so why rock the boat and change everything in one shot? You may find it easier to transition to AVVID if you can also leverage your existing equipment, such as Key System Units (KSU) and PBX equipment. As always, however, issues will arise, and you may want to maintain certain added functionalities. For example, many PBXs can accept dialed digits for the PSTN and international calls. So you may need to configure digit manipulation within your dial peers so you can utilize your current dial plans.

Digit Removal and Prefixes

When a dial string is matched to an outbound POTS dial peer, the terminating router will remove the left-justified digits that were explicit matches for the destination-pattern. The leftover digits would then be forwarded to the telephony device, like the PSTN or the PBX. Sometimes, the telephony interface will need digits removed so they can support the existing dial plan. If this is the case, you can use the command **no digit-strip** in the dial peer configuration. This command will disable the removal of the digits, or you could use the **prefix dial**

peer command, which will prepend digits to the dial string before they are forwarded to the interface. Be aware that these commands only work in POTS dial peers.

Digit Forwarding

You can also limit the number of digits removed from the dialed string. If you use the command **forward-digits {number | all}** on outbound POTS dial peers, the terminating router will not remove the specified digits and will forward them. You can either specify the number of digits that should be forwarded (even if they were explicit matches), or you can use the all switch causing all digits to be forwarded.

Number Expansion

Many larger offices use extension numbers to dial internally between users, instead of the entire E.164 telephone number. Extensions can be defined as a destination-pattern for a dial peer. This way the router will recognize the extension number and will be able to translate it into the E.164 number; that is, if the **num-exp** command has been implemented.

This will enable the router to prepend the digits you define before it passes them to the remote telephony device. This will reduce the total number of digits that must be dialed to complete a call to reach a user at a remote office location. Number expansion is similar to implementing a prefix (discussed earlier), but number expansion is applied to all dial peers, not just those defined.

An example of number expansion would be an office where you would dial the last four digits of the E.164 address to reach someone within the company. In this instance, the complete telephone number may be 747-3637, but internal users would only have to dial 3637 to reach the particular user. All users located at this office have the same first four digits (7473). With this information, you cold configure the dial peers destination-patterns using each extension number, and use number expansion to prepend those first four digits to the extension. The router configurations would look like this:

```
num-exp 3… 7473…
dial peer voice 4 pots
destination pattern 3637
```

Configuring & Implementing…

Digit Manipulation for the Cisco CallManager

You can only apply digit manipulation to route patterns for outbound calls only. This is because the digits need to be sent to the route list plus the route groups. Individual route groups can have specific digit changes for the same route pattern. You usually see this where a dialed number needs to have different modifications like when devices need to dial seven digits to reach a remote office that has a four digit internal dial plan. This often happens when you have a call that cannot be completed through the WAN and needs to be routed though the PSTN. What would occur is the Cisco CallManager would prepend the first three digits onto the dial string. A route pattern can be associated with only one route list.

Route List

A route list is used to route a call. It is configured to map the routes of a call to one or more route groups, which basically act as trunk groups. The route list will then forward the call to the route group based on some predefined preferences. For example, the main (primary) route group may be configured to route calls based on cost and metrics, whereas the secondary route may be configured to only be used in instances where the primary circuit is unavailable, like in an all-trunks-busy condition when there isn't enough bandwidth to admit or complete a call.

Route Groups

In order to control telephony devices like gateways, you create route groups. These gateways can be created using H.323, MGCP, or Skinny Gateway Protocol. End telephony devices that would use H.323 would be programs such as Microsoft NetMeeting and the Cisco CallManager Remote Connections that act as H.323 Gateways. In this setup, the route group can connect to one or more devices, and is able to select between these devices based on preference. In this instance, the route groups can direct all calls destined to the primary device to the secondary device if the main device is not available. Again, this can be considered a trunk group.

You can also point one or more route lists to the same route group. All the devices within this route group have the same characteristics, like path and dial string changes. There is also prioritization, as the string manipulation in a route group overrides the changes of a route pattern.

Telephony Devices

Any IP end device that can be entered into a route group can be considered a telephony device. For example, a device that is configured to use H.323 gateways, such as an IP SoftPhones or Microsoft NetMeeting can be considered a telephony device.

The route pattern dialing structures are usually used to connect IP phone calls destined for external gateways or external Cisco CallManagers using H.323. What this allows for is the ability to use alternate paths if the primary is unable to accept or admit calls. For example, intraoffice calls that use WAN connections as the primary path and the PSTN as the secondary path can choose the secondary path to complete the call if the WAN is saturated. On the other hand, devices that reside on the same Cisco CallManager are unable to use alternate routes, so if there is a problem within the LAN, the phones are unable to reroute to the PSTN to complete the call.

Configuring & Implementing…

The Usage of H.323 Gateways

A device that is "gateway controlled" will need to successfully query the gatekeeper in order to gain admission. The CODEC region should be set to handle the correct CODEC and compression technique. It is allowable to share H.323 gateways between multiple inbound and outbound calls. Gateways that are implemented with Skinny Gateway Protocol and MGCP are only allowed within one Cisco CallManager cluster.

Digit Translation Tables

The ability to manipulate dialed digits is supported within Cisco Call Manager. What this allows for is the manipulation of not only the digits themselves, but also the number of digits within the string. This is most commonly seen in the

directing of calls that have no directly defined destination or DID number. These calls are usually forwarded to an attendant or voice mail. As an example, if your office uses the DID range of 0000 to 0999 and you want calls to be forwarded to the front desk, which is defined as 0001, you can create a translation table of 0XXX with a translation mask of 0001. This will direct calls to the front desk destination. Note: This is for DID numbers that are not defined. This also works for hunt groups, and can be used for internal (also called on-net if within the network) and external (also known as off-net) calls as well as for inbound and outbound calls.

Longest Match Translation

Cisco CallManager is also able to handle the longest match criteria with the implementation of wildcard masks. For example, if there is a phone with a DID located within the 0000–0999 range, the Cisco CallManager will direct the call to that specific phone. In instances where there is no matching extension, then the call will be matched against the translation table, and (using the previous example) be routed to the front desk at 0001.

Digits can also be manipulated within the route pattern configuration using Called/Calling Party Transformation, a method which allows for three types of translations to occur within the called-number. These are:

- The removal of digits from the dialed string
- Application of the Transformation Mask to the called-party
- The ability to prefix digits to the dialed string

These translations can be helpful in companies that either have a lot of unused numbers or companies have multiple numbers. For example, a Cisco CallManager may have a defined route pattern of 8XXXX to route a call to another company office. The number being called is 0000, and needs to go through the PBX. The route pattern has a called party number of 536XXX and a calling party number of 15108XXXX. The calling party information mask of 1510 will be prepended to the calling-number. The access code (the number 8) will be discarded from the dial string, and the digits 536 will be prefixed to the number. All this will happen in that order so it can be properly translated to the internal calling-number of 1510536XXXX so it can be routed by the internal PBX.

The other way to do this is to use the called-party transformation mask of 536XXXX. The drawback to this is that the calling party transformation mask

only applies to the calling-numbers, and the other transformation masks will only apply to called-numbers. As noted earlier, the order of precedence for the Cisco CallManager will be to remove digits from the dialed string, then apply the transformation mask to the called-party, and afterward prefix digits to the dialed string.

Configuring & Implementing…

Options for External Calls Using Route Patterns

As discussed earlier, Cisco CallManager will wait ten seconds before assuming dialing is completed. There are two options that can be added to route patterns destined for outside North America through the PSTN. The more common of the two is dialing the number zero (0).

To configure this option, you could add the statement:

```
Route Pattern = 0.!
```

0. is necessary to access the PSTN, while ! is the wildcard that represents a digit (or number of digits). With this setup, the Cisco CallManager still waits ten seconds to see if any more digits are dialed. If none follow, the Cisco CallManager assumes the dialing is complete and routes the call.

There is also the second option. This configuration instructs users to dial a pound sign (#) to end the dial string so the call can be placed immediately. The drawback is that you are expecting the user to listen to the instruction and change their existing dialing habits. As you know, people aren't always happy with change, especially if they are used to something easier (or that they are familiar with).

```
Route Pattern = 0.!#
```

0. is the code necessary to access the PSTN, while ! is the wildcard that represents a digit (or number of digits). With this setup, the Cisco CallManager will still wait ten seconds to see if any more digits are dialed. If none follow, the Cisco CallManager assumes dialing is complete and routes the call. The **#** (pound) is the end character. When the user dials the pound key, the Cisco CallManager terminates the dialing string and immediately routes the call.

Fixed-Length Dial Peers versus Variable-Length Dial Peers

When considering how to implement your Voice over network, you need to think about the number of digits the router will be dealing with. If you only have fixed-length dialing where users apply four or five-digit dialing to connect to other office phones, the creation of dial plans is really quite simple. You need to know the destination patterns used and build the dial peers based on destination patterns.

On the other hand, some users will need to have full dialing privileges for all their calling needs. When this is the case, you need to implement variable-length dialing plans, something which is bit more complicated. When unsure about the dialing habits of office users, you are generally left with two options:

- You could create a dial plan that includes all possible prefixes and wild cards to ensure all calls are routed (not fun).
- You could implement variable-length dial peers. The router or Cisco CallManager will then collect the dialed string digits and route them based on pattern matching (highly recommended).

Remember that fixed-length peers are exactly that, fixed length. They will always have the same number of digits associated with them whether they are wildcard digits or just dialed digits. For example, if you only configure your router or Cisco CallManager for fixed-length dial plans, the digits received by the router (or Cisco CallManager) must have the appropriate number of dialed digits. If you set up the router to accept ten-digit calls, the router will only connect once all ten digits are dialed. If in this scenario you set up a static area code along with seven digits, and the user doesn't dial that area code, the call will not be able to complete because it does not match the dial peer.

Variable-length dial plans allow for the router or Cisco CallManager to receive inconsistent dialed digits and compare them to its routing table. It can do this through the configuration of several options. For example there can be the inclusion of the command **destination-pattern** with its options. The following configuration of a variable-length dial peer will hopefully give you some idea of what we're talking about, and the explanation that follows should illuminate the configuration.

```
dial-peer voice 1 voip       \\Sets the dial peer as VoIP
destination-pattern 9T        \\dial peer must be matched when the
                              \\router receives the Number 9 + any
```

```
                                    \\digits, or the call will terminate
session target ip4:10.1.100.1       \\When the dial peer is matched, it will
                                     \\ setup a call to 10.1.100.1
```

Several characters, used as switches, can be inserted into the **destination-pattern** command. The preceding configuration uses the "T" switch, which is a timeout character. You could also configure a termination character defined with the command **dial peer terminator <character #*[0-9]>**. I prefer to use "#" for termination, but you can choose other characters. Keep in mind, though, that this command can only be found on routers that are voice-enabled. To enable the termination character, you can do any of the following:

- Use the "#" as termination character that can be sent from the Telephony device, thus making it like a Cell phone send key. So the router would receive the "#" character and know that it need to send all of the characters that were dialed before the "#" key.
- When configuring the voice-port, you can add the command "timeouts interdigit" and define the amount of time that router or Cisco CallManager will wait between dialed digits (normally set to 10 seconds by default) before sending the digits. You may want to configure this for a smaller interval, as many times users will become inpatient with this long a wait.

When you install Cisco CallManager within North America, you can use the "@" character with the creation of route patterns to create variable length dial plans. This way the user can dial a seven-digit local number, or ten- (area code + number) and eleven- (1 + area code + number) digit numbers to call long distance. When the number reaches the last dialed digit, the call will immediately be placed. The "@" character will not work outside North America, though.

In order to construct variable-length dial plans in the past, you needed to configure the Cisco CallManager with a router pattern that consisted of **0.!** within the setup. By setting up the wildcard, the Cisco CallManager would then be able to utilize variable-length dial plans, but it also needed to use the timeout after the last digit before it would place a call to the gateway. The alternative was to create variable-length dial plans for the entire national calling-number scheme. A lot of support issues needed to be addressed to make this feasible, but it allowed a myriad of calling features and offered users a minimal wait.

For international calls, you will need to implement the wildcard setup, as North American systems are not designed to match foreign exchanges.

What Is Two-Stage Dialing?

When a voice call is destined for the network, the router placing the call collects all the dialed digits. It then takes these digits and filters them through the dial peers to see if there is a match. Once a match is found, the router then immediately places the call by forwarding the dialed string. Once the call is forwarded, the router no longer collects digits for that session and they are dropped. Digits and wildcards used in the destination pattern choose how many digits the router collects before it tries to filter them through the dial peers.

Matching Variable-Length Dial Peers

Routers are configured by default to match variable-length dial peers. As long as the digits dialed match the pattern on the dial peer, it will continue to filter. Once you are processing digits beyond the matching point, however, the router will ignore them during the filtering process. For example, the dial string for information, 5551212, would be properly matched with the following dial peers:

```
dial-peer voice 1 voip
destination-pattern 555
session target ipv4:10.1.100.1
dial-peer voice 2 voip
destination-pattern 5551212
session target ipv4:10.1.100.2
```

In order to disable the matching of variable-length dial peers, you would add the $ character at the end of the destination-pattern. The $ character will stop the dial peer from matching the digits that would come after it, even if they were able to be processed by another destination-pattern, as in the following example:

```
dial-peer voice 1 voip
destination-pattern 555$
session target ipv4:10.1.100.1
dial-peer voice 2 voip
destination-pattern 5551212
session target ipv4:10.1.100.2
```

With the $ at the end of the destination pattern, the dial peer for 5551212 would not be matched. The pattern would only match up to the 555 configured for dial peer 1.

As noted earlier, two-stage dialing collects digits that are dialed. It actually collects them one by one and will attempt to match a dial peer after each digit is dialed and processed. Once a match is found, the call will be routed. So, dialing 5551212 and using the following configuration:

```
dial-peer voice 1 voip
destination-pattern 555
session target ipv4:10.1.100.1
dial-peer voice 2 voip
destination-pattern 5551212
session target ipv4:10.1.100.2
```

you would see that the router would match the digits immediately to dial peer 1 and route the call.

In order for the digits to match the second dial peer, you would need to use the timeout character (T) at the end of the destination pattern, in this case 555. This would allow the digits a time limit with which to dial all numbers, and that would allow the pattern to be matched to the best fit. This configuration would look something like this:

```
dial-peer voice 1 voip
destination-pattern 555T
session target ipv4:10.1.100.1
dial-peer voice 2 voip
destination-pattern 5551212T
session target ipv4:10.1.100.2
```

Be aware that the router will also select dial peers based on whether the call is inbound or outbound.

Creation of Calling Restrictions and Configuration of Dial Plan Groups

Within Cisco CallManager, you can create calling restrictions on a per telephony device, or create closed dial plan groups (as long as they fall within the same Cisco CallManager). What this means is that users that reside within the same Cisco CallManager can be grouped together with the same calling restrictions and dial plans. For example, if you have development teams that only need to talk to each other, you can restrict their dial-plans to within the group, or limit their

ability to call long distance. Within the same Cisco CallManager, you may have Accounting or Human Resources that need to make more long distance calls, so you create calling communities based on their need.

These different communities are able to operate independently and can all share the same gateway since they have overlapping dial-plans. You will find this more useful in sites linked across WAN connections if they all share a central Cisco CallManager as the call processing area. This also allows for the usage of partitions and calling search space within the organization.

Partitioning with Cisco CallManager

So what is a partition? A partition is a group of telephony devices that have similar reach ability. These devices are composed of route patterns, IP SoftPhones, directory numbers, and so on. When creating partitions spaces, it's a good idea to group together those with similar characteristics and give them a name that reflects those qualities. For example, if you have your System Engineers in building A**, North** then you should create a group name something like SE-AN.

Creating a Calling Search Space

What is a calling search space? It is a list of partitions that can be accessed by users so they can place a call. These calling search spaces are only allocated to telephony devices that can start calls. Once implemented, it is simple to create and use dialing restrictions because users are only allowed to dial those partitions in the calling search space they are assigned to. If the user tries to call outside the allowed partitions, they receive a busy signal.

For all intents and purposes, the calling search space is what allows callers to complete connections for their calls. You would often use this configuration when setting up office call policy. For example, when you set up office phones, you often allow them unrestricted dialing abilities. Lobby phones, on the other hand, can usually only call other phones located in the office. To establish these criteria, you must create a partition for the office users, in this case SE-Users (see Table 9.6). All calls destined for the PSTN would have the route pattern 9, and those calls would be placed within the SE-PSTN allocated partition. Two calling search spaces would then need to be created to represent the two sets of dialing characteristics (see Table 9.7).

Table 9.6 The Assignment of Partitions

Partition Name	Devices Designated to Partition
SE-Users	All office telephony devices
SE-PSTN	All devices with routes destined for the PSTN

Table 9.7 The Assignment of Calling Search Space

Calling Search Space	Partitions	Devices Assigned
Unrestricted	SE-Users SE-PSTN	All telephony devices that can make calls
SE-Internal	SE-Users	Telephony devices that cannot call outside the local office

One of these calling search spaces would be labeled Unrestricted (to denote the lack of restrictions on the calling device). This calling search space would then have SE-Users and SE-PSTN associated with it. The second calling search space (called SE-Internal) would then have only SE-Users associated with it.

Office users in the Unrestricted calling search space will be allowed to dial anywhere, while telephony devices associated with the SE-Internal calling search space will only be allowed to call internally.

From this basic configuration, you could add all sorts of calling features, depending on the needs of your office. These include:

- Limiting telephony devices to intrasite (local office) calling
- Limiting telephony devices to intrasite calling, with emergency calling ability (emergency calling is required for most, if not all, office configurations)
- Intrasite and intersite (external offices) calling
- Intrasite and intersite calling, with emergency calling capability
- Intrasite, intersite, emergency, and local PSTN calling
- Intrasite, intersite, emergency, and national PSTN calling
- Unrestricted calling (includes all the preceding, plus international calling)

These partitions and calling search spaces can allow autonomous dial ranges on a partition basis. Extension and access codes located within different partitions

can have overlapping number schemes, and still work independently of each other. This is usually seen in the implementation of a centralized call processing system. In this example, all sites that use the same Cisco CallManager can dial the number 9 to access the PSTN, even if they are located on different WAN segments.

When using a centralized Cisco CallManager for call processing, certain conditions apply to overlapping users and extensions located at other sites. These include:

- Overlapping internal dial plans are supported if there is an implementation of, or need for, voice mail on those extensions. This prevents issues with the Cisco CallManager sending calls to voice mail and having to decide which partition the call is destined for. The Cisco CallManager is not designed to be intuitive, so a call directed to ext. 3637 in Seattle cannot be distinguished from a call directed to ext. 3637 in San Francisco. Voice mail requires unique extensions for identification.
- If you do not require voice mail, you can have multiple sites with the same extension. These extensions can be reached via:
  - The PSTN, dial the area code (if necessary), local access (exchange) code (747), followed by the full directory number (3637).
  - The WAN, through the implementation of translation tables. The tables can allow prepending of a unique code (sometimes referred to as a steering code) to occur on extensions that overlap. This steering code is then removed from the call when the destination is reached.

Guidelines for the Design and Implementation of Dial Plans

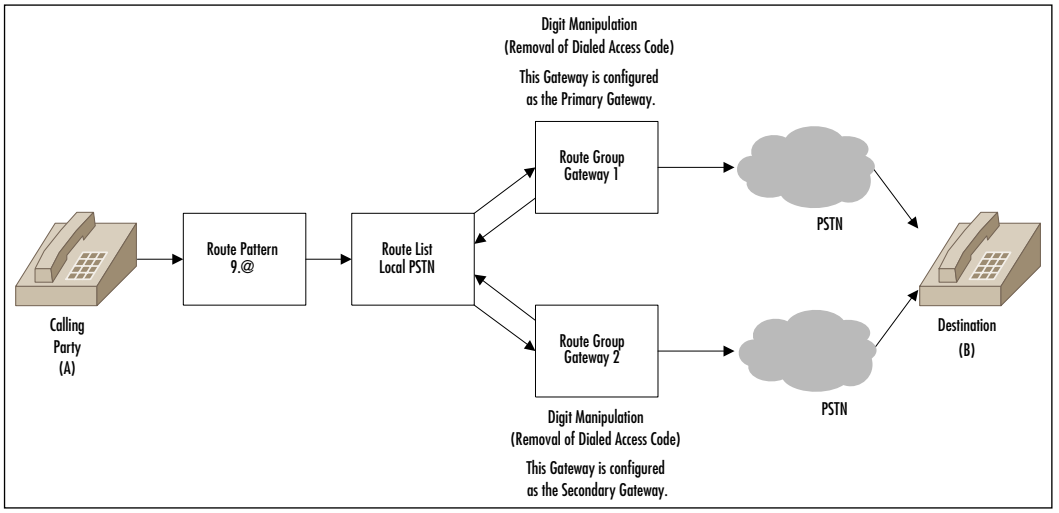
As with any project, its complexity will depend on the number of variables factored in. Dial plan complexity can vary, based on any number of configuration choices, such as the total amount of paths a call can be sent through. What I will do in the following section is try to give you an idea of what to expect with some of the usual dial plan implementations.

Setting Up Single-Site Campuses

In many instances, you will implement AVVID-based solutions in a single site configuration. These are the implementations that only have one office and no

WAN connections to external sites. When configuring these types of sites, you will often implement a simple dial plan that can provide intraoffice calling (with four or five digits depending on the site), as well as connections to the PSTN (usually by dialing a 9). Long distance would also be handled by the PSTN, with the dialing party using a 9, then a 1, followed by the area code before dialing the seven-digit number. If you plan to use multiple carries for your PSTN, you may have a scenario that flows like that in Figure 9.4.

Figure 9.4 Cisco CallManager Flow Chart for Single Campuses



The dial plan is set up to use one route pattern. The 9.@ is the configuration pattern that signifies the 9 as the access code to connect to the PSTN. The @ is required to configure the dialing plan as the North American standard (E.164). The "." is used by the Cisco CallManager to tell it which digits are considered after the access code. This needs to be configured to be sure to remove the correct digits (the digits located on the left of the dot).

The route pattern will also allow the dialing of 911 for emergency services. The route group is configured to remove the access code (9) from the dialed string so the call can be properly routed through the PSTN.

You would often see the multiple PSTN setup for redundancy. This way, if one PSTN becomes unavailable, or the gateway connected to the PSTN does not function, the Cisco CallManager will route the call through the secondary gateway.

When configuring a Cisco IOS H.323 gateway, try to minimize the number of entries. For the most part, the dial plan configurations should occur at the Cisco CallManager. This adds to the efficiency of the router. You could also configure these gateways to use the Skinny Gateway Protocol or MGCP, but you will more commonly use the H.323-based gateways.

```
dial-peer voice 1 voip
codec g711ulaw                  \\This states that the Dial peer for
                                \\ all incoming calls from PSTN to
                                \\ Cisco CallManager's IP address must
                                \\be G.711
dtmf-relay h245-alphanumeric
destination-pattern 9....
session target ipv4:10.1.100.1  \\This is the Cisco CallManager's IP
                                \\address
!
dial-peer voice 2 pots          \\This is the Dial peer for all 7-digit
                                \\outgoing PSTN numbers
destination-pattern......
port 1/0:1
!
dial-peer voice 3 pots          \\This is the Dial peer for all 10-
                                \\digit outgoing PSTN numbers
destination-pattern 1.......
prefix 1
port 1/0:1
!
dial-peer voice 4 pots          \\This is the Dial peer for 911
                                \\services
destination-pattern 911
prefix 911
port 1/0:1
```

With this configuration, the Cisco CallManager assumes the 1 + 10 digit dial string is required to make long distance calls through the PSTN, and that seven-digit calling would use the PSTN for local calls. Even though the addition of the

9.@ (as discussed earlier) includes the ability to dial 911 for emergencies, the Cisco IOS gateway requires an entry for the dial peer. You could also add dial peers for 411 if you want that service to be available (this is also handled by the 9.@).

NOTE

Voice over IP points to layer three (routed) addresses, so it looks like IP packet-based traffic as it traverses the network. POTS traffic (and VoFR) is based on layer two, so it is treated as a voice call all the way to completion.

Design Considerations for the Creation of a Dial Plan

For this example, we will discuss the implementation of a national dial plan for a location that resides within the United States. This methodology can be used to create other plans anywhere in the world if you implement the proper dialing schema.

Designing a Dial Plan to Meet Your Needs

Designing a dial plan that meets your needs sounds pretty fundamental, but what does it mean? When you implement AVVID, you should work under the assumption that the less complex it is, the better. Find out what is used on a normal (daily) basis, and what features are seldom used. With these answers, you can create a plan that meets the needs of the client.

If you are setting up a branch office, you will probably need to implement a system similar to this. Company X would like to set up AVVID within their regional offices. All branch offices will have several levels of call barring that allow local calls (those calls located within the local exchange only), some that allow long distance calls, and some that allow international calling. For ease of demonstration, we will create dial plans that allow for a greater level of granularity then you might encounter in the real world. By creating a high level of distinction, you will be able to filter numbers using the local dialing prefix from all other number combinations. This will help place these router patterns into separate partitions and calling search spaces (as discussed earlier). This setup allows for the control of end telephony devices and their ability for outdial access.

NOTE

You will need to alter the dial plan to fit the local numbers where the CallManager is located. The [ ] wildcards will allow you to specify a range of numbers, which should reduce the overall number of route patterns that are comparable.

The United States (and North American) standard for dialing is based on seven digits for the local exchange area, three digits for the area code, and a leading one (1) for long distance calls (1 + 425 + (555-5555)). Table 9.8 shows a basic dial plan.

NOTE

Table 9.8 is not an exhaustive list of all possible call combinations. It is quite possible there are other numbers that aren't listed, so please investigate the particular dial plans associated with your location. Phone books often have lists of area code and informational/service numbers.

Table 9.8 Sample Dial Plan

Route	Pattern	Comments
9.911	No Pattern	Standard number for emergency calls
9.411	No Pattern	Standard number for Information
0.5551212	No Pattern	Standard number for Information
9.1[800]XXXXXXX	1800XXXXXXX	Toll-Free Call
9.1[866]XXXXXXX	1866XXXXXXX	Toll-Free Call
9.1[877]XXXXXXX	1877XXXXXXX	Toll-Free Call
9.555XXXX	No Pattern	Local Exchange Numbers—seven-digit numbers
00!	No Pattern	International—uses interdigit default timeout (ten seconds)
00!#	No Pattern	International—uses # to signal end of dial character

Configuring the Dial Plan within the Cisco CallManager

Configuring the dial plan within CallManager is a fairly straightforward process, which can be distilled down to these easy steps:

1. Enter an access code of **9.** as the access code delimiter. You can add the route pattern digits, wildcard matches, or the "@" character for North America.
2. Ensure that the **Route this pattern** and **Provide secondary dial tone** options are set.
3. Point the route pattern to a gateway device (**H.323**, **MGCP**, **Skinny Gateway Protocol**, **SIP**).
4. If the gateway device is MGCP or Skinny Gateway Protocol, you must make sure the access code is discarded. To do this, set discard digits to **<pre-dot>** under Called Party Transformations.
5. If the gateway device is a Cisco IOS-based H.323 gateway, then the access code will need to be passed along with the called digits. To do this, set discard digits to **<none>** under Called Party Transformations.
6. Create the route pattern in the database.

Verifying the Dial Plan is Correct

Based on the assumption that this will be a Cisco IOS-based H.323 gateway, you would then point the router POTS dial peer to the PSTN port (or ports) and use a destination-pattern of "9" to match the leading digit that will come from the Cisco CallManager. The match on the "9" will make the dial peer remove the 9, so the rest of the number is passed.

```
dial-peer voice 100 pots
 direct-inward-dial       \\Creates the DID for incoming calls
!
 destination-pattern 9    \\Removes the 9 when a call is placed
!
 port 1/0:15               \\This will direct the call to the PRI port
                           \\ 1/0
!
!
```

You will not need multiple POTS dial peers if you are not setting up a hunt group that will use multiple POTS. If you were to set up multiple POTS for hunt groups, and you wanted the calls to cycle through several configured voice ports, it might look like this:

```
!
dial-peer voice 100 pots
 destination-pattern 9
 port 1/0/0
 !
dial-peer voice 101 pots
 destination-pattern 9
 port 1/0/1
!
```

Creating a Dial Plan for a Multisite Organization

For the creation of a Dial Plan for a multisite WAN, you must have sufficient resources to make it function properly. If you don't have the proper link bandwidth, the call will always route over the PSTN, therefore negating the benefits that multisite WAN connections are supposed to give you. If the system can route calls over the WAN then you will be able to bundle the costs of long distance with your data connection, and hopefully maximize your investment. When you implement the dial plan with the proper gatekeeper call admission control mechanism, the dial plan will decide when to place calls across the WAN and what it will do if the gatekeeper will not accept the call.

The Role and Configuration of a Cisco CallManager and Gatekeeper

By implementing H.323 gatekeepers for admission control, you will be able to control the number of calls allowed to and from specific areas. This will assist you in the management of bandwidth and resources for your sites and overall infrastructure. The Cisco CallManager uses the gatekeeper to perform admission control, especially in infrastructures that use hub and spoke architecture for network centralization.

The gatekeeper will also assist with address resolution. The Cisco CallManager will know what ranges of extensions are within its control and is able to route the calls to the proper destination. If the extension is not local to the Cisco CallManager, it goes to the gatekeeper for the address of another Cisco CallManager that can direct the call to its destination (hopefully).

The reason this occurs is because each Cisco CallManager registers itself with the gatekeeper, which is configured with the address range of each Cisco CallManager (or cluster). This helps hold the dial plans each Cisco CallManager will need to set up. There are three types of dial plan deployment models for destination resolution of calls from within the cluster. They are:

- The Cisco CallManager Model
- The Gatekeeper Model
- The Hybrid Model

The Cisco CallManager Model

This dial plan model requires that all Cisco CallManagers located within a cluster be connected through an intercluster trunk with a route pattern for each of the other clusters within the domain. This creates quite a bit of overhead as you add new Cisco CallManager clusters since you will have to reconfigure the Cisco CallManager every time a new cluster is added or a dial plan changes. With this deployment, the gatekeeper handles call admission control but not dial plan resolution. This setup is pretty close to what you might think of as a normal PBX. The Cisco CallManager will need two routes configured for each destination, one for the WAN and one for the PSTN. Calls are then automatically routed over the PSTN if there are insufficient resources on the WAN.

The Gatekeeper Model

This dial plan model helps clean up the overhead inherent within the Cisco CallManager model. This is because the Cisco CallManager only needs to maintain one intercluster trunk known as the *anonymous device*. This device is like a point-to-multipoint connection in frame relay, as the Cisco CallManagers don't need to be fully meshed. In this set up, the gatekeeper is able to use the anonymous device to route calls through the network to the correct Cisco CallManager (or cluster).

You can set this model up for exclusionary clusters. What this means is that there is no automatic overflow for calls destined for the WAN which are unable to complete due to insufficient resources. In this configuration, you need two route patterns, one to the WAN and one to the PSTN. When users are unable to connect over the WAN, they need to dial the PSTN number to access the other site. The dial plans are configured and executed within the gatekeeper, so when you add new Cisco CallManagers, or change the dial plan, you only need to configure the gatekeeper.

The Hybrid Model

This type of configuration incorporates the best of both worlds. There is automatic overflow to the PSTN for calls destined for the WAN which are unable to allocate sufficient resources. It only needs one anonymous device for each Cisco CallManager (cluster), thus minimizing the overhead of having to mesh the Cisco CallManagers. It does require two routes for each destination, however, one to the WAN and one to the PSTN. The drawback is you need to configure the dial plan on the gatekeeper and the Cisco CallManager.

Configuration of a Cisco CallManager Cluster

For every gatekeeper located within your domain, you must configure an intercluster CODEC you would like to use, and along with this enable the anonymous device. When that is complete, you will need to configure the router pattern to allow calls between clusters. You do this by selecting a CODEC for all intercluster calls, defining the region the gatekeeper and cluster are located in, and selecting the appropriate compression rate.

When the Cisco CallManager registers with the gatekeeper, it identifies itself as a VoIP gateway that has a technology prefix of Voice. You can manually set this configuration by going to the Service tab, scrolling down to the Service Parameters, and from there updating the Cisco CallManager service parameter GateKeeperSupportedPrefix configuration. Figures 9.5, 9.6, and 9.7 illustrate the basic configuration of a Cisco CallManger.

NOTE

The GateKeeperSupportedPrefix is hidden by default. Refer to your Cisco documentation regarding how to access it.

Figure 9.5 Configuring a Cisco CallManager Device Pool

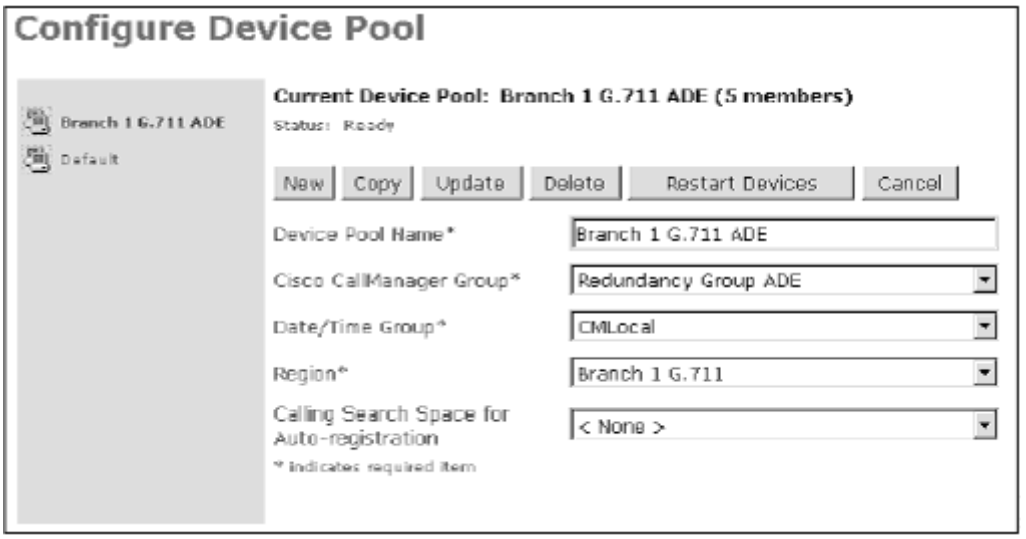


Figure 9.6 Configuration of a Region with Cisco CallManager

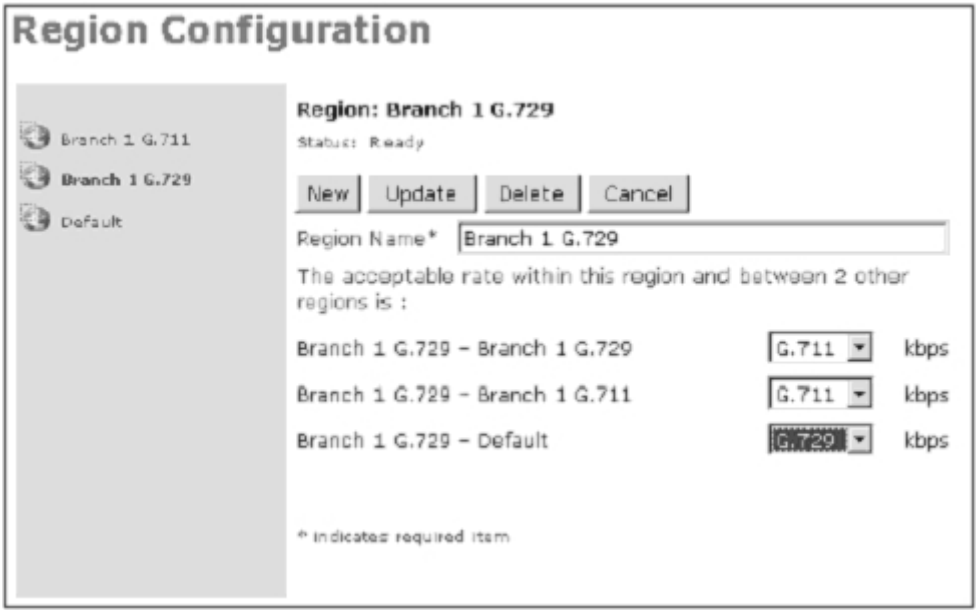
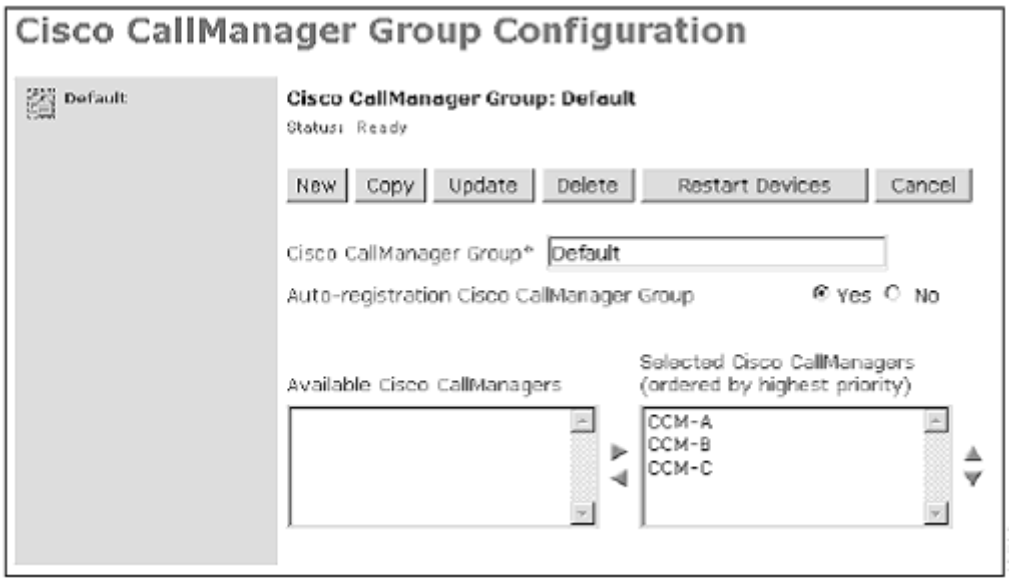


Figure 9.7 Configuration of a Group with Cisco CallManager



Gatekeeper Configuration

When configuring the Cisco CallManager gatekeeper, you are required to enter a zone. Each Cisco CallManager will register with that zone: its zone prefix (this is the directory number ranges), the bandwidth allowed for each call admission, and the technology prefix for voice-enabled devices. Cisco CallManager will need the gatekeeper to explicitly specify the IP address of the Cisco CallManager within a single zone, and then you must disable the registration of all other IP address ranges so it only exists within that zone. Cisco CallManager will register with the gatekeeper using its IP address as its H.323 ID.

Video Dial Plan Architecture

Technically, there is a lot of overlap between the configuration of voice and video on the network. The following sections are written to explain some of the differences, but keep in mind that topics discussed earlier within the chapter still need to be considered when implementing a video dial plan.

When corporate videoconferencing was first introduced in the 1980s, people saw it as a way to help connect people located in different cities so they could communicate more effectively and efficiently, and at a greater savings in cost.

With this in mind, there came the creation of standards that would allow these remote locations to connect to each other. These first-generation solutions were based upon the International Telecommunications Union (ITU) H.320 standards defining Integrated Switched Digital Network (ISDN) connection-based videoconferencing.

Not long after this, second-generation solutions were created, bringing videoconferencing to computer desktops. The main drawback to this more refined technology was that it was still dependent upon ISDN and expensive CODEC devices, so it wasn't feasible to offer it to most users for normal business needs. In the mid-1990s, the creation of the third-generation, LAN-based solutions, became more prevalent in many organizations. Accessible desktop videoconferencing applications multiplied, but until recently, these had remained proprietary, and quite often, very expensive.

Because of compatibility issues between different device and software solutions, widespread deployment up to this time has been severely limited. With recent changes in world travel, however, you may begin to see videoconferencing become more prevalent in the next few years; a trend that should, in turn, lead to increased attempts at standardization as more companies try to corner the market.

You have to remember that when H.323 and H.324 standards were created, they gave software and hardware vendors the ability to create videoconferencing packages that were more manageable and affordable. This isn't to say they solved all problems are associated with videoconferencing (circuits are not always cheap, and do not necessarily offer the uptime and dedicated bandwidth necessary to support every organization).

The H.323 standard was able to define videoconferencing technologies and enabled multivendor interoperability for the first time. Meanwhile, the H.324 standard was defined as a solution for videoconferencing using POTS lines. When Cisco Systems started to deploy AVVID, they were able to enable end-to-end, global desktop videoconferencing, VoIP, multilocation collaboration, and electronic whiteboard applications.

As more and more applications based upon H.323 become available, and the cost of digital video cameras continues to decline, there will be a more ubiquitous implementation of desktop video conferencing. Also, as companies become more and more economically savvy, you will see many accountants wanting to leverage existing technologies and add one time costs (such as the digital camera) to help with the bottom line and maintain cash flow.

Cisco introduced its Multimedia Conference Manager in February of 2001. It is an H.323-compliant software that enables its users to create policy-based

management for H.323 applications, and is one of the first multifunctional, scalable solutions available for a large market. It allows you to control and manage voice and video connections with much more efficiency given your network resources.

The Cisco Multimedia Conference Manager

The Cisco Multimedia Conference Manager (Cisco MCM) is a specialized Cisco IOS software image that lets network administrators support H.323 applications on their networks without compromising mission-critical traffic from other applications. The Cisco MCM serves two main functions: it acts as a gatekeeper, and as a proxy. The following sections will help you understand what the differences are between these two functions.

Gateway

A gateway is an optional element that can be implemented within the H.323 deployment. It is an endpoint on the LAN that can provide real-time, two-way communication between H.323 terminals or other gateways. It is also capable of using the LAN and other ITU terminals located on the WAN by using H.425 and Q.931 protocols. These gateways are not required if there are no connections to other networks. You will use these gateways when you need to:

- Create connections over analog PSTN terminals
- Create connections with H.320-based terminals across ISDN circuits
- Create connections with H.324-based terminals over a PSTN networks

Gateways can translate between H.323 conferencing endpoints and compliant terminals, H.225 to H.221, or between communication procedures, H.245 to H.242. These gateways can also translate between audio and video CODECs, performing call setup and termination on both sides of a network. If you configure the gateway properly, it can also support H.310, H. 321, H.322, and V.70 standard terminals.

Proxy Gateway

A proxy gateway is a secured connection between H.323 sessions, allowing Cisco Multimedia Conference Manager to contain a proxy as part of its infrastructure so it can provide QoS, traffic shaping, and security and policy management for H.323 traffic across any secured connection. A proxy in this case is very similar to

proxy gateways in other network deployments. In essence, it is a go-between that is able to connect two dissimilar connections, or two connections that need to be separated. It can establish and maintain its connection between these multiple end points, adding to your overall efficiency by offloading some of the features to the proxy gateway.

The H.323 Gatekeeper

The H.323 Gatekeeper is an optional component that provides call control services for H.323 endpoints within your network. You can implement multiple gatekeepers to run within your network, so long as they remain logically separate from the endpoints. There are currently no standards for gatekeeper-to-gatekeeper communications as of yet, so you may want to explore other options before installing multiple gatekeepers within the same segment. Keep in mind, this can cause some implementation issues, and, if implemented within your main LAN/WAN, some network issues as well. To alleviate these problems, you could install terminals, Multipoint Control Units, gateways, or other non-H.323 LAN devices since these can coexist in the same environment. H.323 Gatekeepers provide the following functions:

- **Address translation** Alias addresses can be maintained to transport address translation. The gatekeeper often uses a translation table that is updated by using registration messages (other table update methods are allowed as well, such as manual updates).
- **Admission control** This authorizes access using ARQ/ACF/ARJ/H225.0 messages, based upon call authorization, bandwidth, or other criteria set by the vendor and your predefined settings. You can also set this as a null function to admit all requests without performing filtering.
- **Bandwidth control** The gatekeeper supports BRQ/BRJ/BCF messages based upon bandwidth management. This can also be set as a null function that accepts all requests for bandwidth changes.
- **Zone management** Used for registered terminals, Multipoint Control Units, and gateways.

Multipoint Control Units

So, what is a Multipoint Control Unit (MCU)? An MCU is a device that aids in getting calls out to three or more endpoints in conference type deployments. It works like a meet-me call bridge, and is usually a centralized device that assists in the facilitation of conference sessions for data, video, and/or audio.

Configuring Video Dial Peers

Video dial peers is a feature supported only on the MC3810 Multi-Service Concentrator. What follows is a basic set of commands and how they act within the presented command set. These are very basic, and should get you to minimal running specifications.

1. Specify which slot the Video Dialing Module (VDM) is located in. The keyword *slot* is the location value of the VDM (it will either be 1 or 2). The keyword *port* is the interface. The MC3810 chassis only has one VDM so the value is 0.

   ```
   MC3810(config)# port signal slot port
   ```

2. Define the ATM dial peer for the remote system. Video dial peers will remain until they are explicitly removed. The keyword *tag* identifies the dial peer. It must be unique on the MC3810 and must be a number from 1 to 1000. The keyword *videocodec* identifies the video CODEC associated with the router. The keyword *videoatm* identifies the video CODEC associated with the ATM network.

   ```
   MC3810(config)# dial-peer video tag {videocodec | videoatm}
   ```

3. You must specify the E.164 address that will be associated with the dial peer. The keywords are explained earlier in the chapter with the exception being the + sign, which is not supported on the MC3810.

   ```
   MC3810(config-dial-peer)# destination-pattern [+] string [T]
   ```

4. On the Cisco MC3810, you must configure the ATM session target for the dial peer. The keyword *serial* specifies the interface for the dial peer address. The *atm* keyword identifies the ATM interface number. The *interface* keyword identifies the interface number. The *svc nsap* keyword identifies the SVC (Switched Virtual Circuit) and the nsap (Network Service Access Point). The *nsap-address* is the 40-digit hexadecimal

number for the session target nsap. The keyword *pvc* is the Permanent Virtual Circuit (PVC). The keyword name identifies the target ATM pvc. The keyword *vpi/vci* is the ATM network Virtual Path Identifier (VPI) and the Virtual Channel Identifier (VCI) for this PVC. Lastly, the *vci* keyword identifies the ATM network VCI for this PVC.

```
MC3810(config-dial-peer0# session target {serial | atm} interface
    {svc nsap nsap-address | pvc}{name | vpi/vci | vci}
```

From here, you would repeat these steps on the routers involved in the configuration.

NOTE

By using PVCs to send your data traffic, you would be able to identify a PVC defined on the ATM interface as a session target by using a name or VPI/VCI pair.

Summary

In this chapter, we talked about how a dial plan is, in its most basic form, a system interface for telephony devices that allows users and equipment to connect to each other by using dialing strings. These dial strings can be mapped and routed to a multitude of locations by the controlling system. AVVID implementation uses routers like filters to match dial peers. When a call arrives on a POTS port, the router will match the VoIP dial peer for the outbound call. A dial plan allows you to design your network so it can accommodate data and voice/video within the same infrastructure. A dial plan can be created in many different ways, depending on the needs surrounding each individual deployment.

These plans can include redundant paths, for IP networks and the PSTN, so there will be a path to completion for calls made.

By using Cisco CallManager, you are able to allow for greater growth and functionality within your network because Cisco CallManager was designed to be integrated with IOS gateways to allow for greater functionality while streamlining the installation.

We also saw that within Cisco CallManager, you can create calling restrictions on a per telephony device, or create closed dial plan groups as long as they fall within the same Cisco CallManager. This ensures that users residing within the same Cisco CallManager can be grouped together with the same calling restrictions and dial plans for ease of administration.

We also talked about the implementation of H.323 gatekeepers. These gatekeepers handle admission control so you are able to control the number of calls allowed to and from specific areas. We saw how this would assist you in the management of bandwidth and resources for your sites and overall infrastructure. The gatekeeper can also assist with address resolution. The Cisco CallManager knows what ranges of extensions are within its control and is able to route the calls to the proper destination. If the extension is not local to the Cisco CallManager, it will then go to the gatekeeper for the address of another Cisco CallManager that can direct the call to its destination.

In the LAN, it is becoming more and more feasible to introduce video and other multiservice technologies to better leverage the local area network and its increasing speed. With the implementation of AVVID and the usage of H.323 gateways, video to the desktop is becoming more and more commonplace. By using Cisco AVVID video gateways, internal users can communicate more efficiently and seamlessly.

Solutions Fast Track

What Is a Dial Plan?

- ☑ Configuring dial peers for use is essential when designing and implementing Voice over IP on your network. Dial peers identify the calling source and the destination points so as to define what attributes are assigned to each call.
- ☑ Configuring a dial peer for POTS can help you shape the deployment of your dial peers.
- ☑ By configuring VoIP dial peers, you can enable the router to make outbound calls to other telephony devices located within the network.
- ☑ Dial peers for inbound and outbound calls are used to receive and complete calls. You must remember that the definition of inbound and outbound is based on the perspective of the router. What this means is that a call coming into the router is considered an inbound call and a call originating from the router is considered an outbound call.
- ☑ To associate a dialed string with a specific telephony device, you would use the destination pattern. With it, the dialed string will compare itself to the pattern and then will be routed to the voice port or the session target (discussed later) voice network dial peer. If the call is an outbound call, the destination pattern could also be used to filter the digits that will be forwarded by the router to the telephony device or the PSTN. A destination pattern must be configured for each and every POTS and VoIP dial peer configured on the router.
- ☑ The session target is the IP address of the router to which the call will be directed once the dial peer is matched.
- ☑ Route patterns (on-net) allow you to connect to multiple sites across a WAN with connections like frame or dedicated circuits using available network resources.
- ☑ With Cisco CallManager, you are able to create route patterns that allow you to route calls that differentiate between local calls and long distance calls.

Cisco CallManager Dial Plans

- ☑ By using Cisco CallManager, you can allow for greater growth and functionality within your network because it was designed to be integrated with IOS gateways.
- ☑ The creation of dial plans for internal calls to IP phones are registered within a Cisco CallManager cluster.
- ☑ External calls use a route pattern to direct off-network calls to a PSTN gateway. Route patterns can also be used if there are Cisco CallManagers located on a WAN-connected network.
- ☑ A route pattern is the addressing method that identifies the dialed number and uses route lists and route group configurations to determine the route for call completion.
- ☑ Digit manipulation involves digit removal and prefixes, digit forwarding, and number expansion.
- ☑ Route lists are configured to map the routes of a call to one or more route groups.
- ☑ Route groups allow you to control telephony devices.
- ☑ Telephony devices are any devices capable of being connected to a route group.
- ☑ the digit translation table manipulates dialed digits and is supported within Cisco Call Manager
- ☑ Fixed-length dial peers versus Variable-length dial peers—This will help you to decide what to use in your network.
- ☑ Two-stage dialing occurs when a voice call is destined for the network, and the router placing the call collects all of the dialed digits.

Creation of Calling Restrictions and Configuration of Dial Plan Groups

- ☑ Within Cisco CallManager, you can create calling restrictions per each telephony device, or create closed dial plan groups (as long as they fall within the same Cisco CallManager). What this means is that users residing within the same Cisco CallManager can be grouped together

with the same calling restrictions and dial plans. For example if you have development teams that need to talk to only each other, you can restrict their dial plans to within the group, or limit their ability to call long distance.

- ☑ A partition is a group of telephony devices that have similar reach ability. These devices are composed of route patterns, IP SoftPhones, directory numbers, and so on.
- ☑ A calling search space is a list of partitions that can be accessed by users in order to place a call. These calling search spaces are only allocated to telephony devices that can start calls. With these calling search spaces implemented, it is simple to create and use dialing restrictions, because users are only allowed to dial those partitions in the calling search space they are assigned to. If the user tries to call outside the allowed partitions, they will receive a busy signal.
- ☑ The combination of partitions and calling search spaces can allow autonomous dial ranges on a partition basis. Extension and access codes located within different partitions can have overlapping number schemes, and will still work independently of each other. This is usually seen in the implementation of a centralized call processing system. In this example, all sites that use the same Cisco CallManager can dial the number 9 to access the PSTN, even if they are located on different WAN segments.

Guidelines for the Design and Implementation of Dial Plans

- ☑ As with any project, its complexity will depend on the number of variables factored in. Dial plan complexity can vary, based on any number of configuration choices, such as the total amount of paths a call can be sent through.
- ☑ When configuring single-site campuses, you will often implement a simple dial plan that can provide intraoffice calling (with four or five digits depending on the site) and connections to the PSTN (usually by dialing a 9). Long distance would also be handled by the PSTN with the dialing party using a 9, then a 1, and the area code before dialing the seven-digit number.

- ☑ When you go to implement AVVID, you should work under the assumption that the less complex it is, the better. Find out what is used on a normal (daily) basis, and what features are seldom used. With these answers, you can create a plan that meets the needs of the client.
- ☑ Based on the assumption that this will be a Cisco IOS-based H.323 gateway, you would then point the router POTS dial peer to the PSTN port (or ports) and use a destination pattern of "9" to match the leading digit that will come from the Cisco CallManager. The match on the "9" will make the dial peer remove the 9, so the rest of the number is passed.
- ☑ When creating a dial plan for a multisite WAN, you must have sufficient resources to make it function properly. If you don't have the proper link bandwidth, the call will always route over the PSTN, negating the benefits that multisite WAN connections are supposed to give you.

The Role and Configuration of a Cisco CallManager and Gatekeeper

- ☑ By implementing H.323 gatekeepers for admission control, you can control the number of calls allowed to and from specific areas. This will assist you in the management of bandwidth and resources for your sites and overall infrastructure. The Cisco CallManager uses the gatekeeper to perform admission control, especially in infrastructures that use hub and spoke architecture for network centralization.
- ☑ The Cisco Call Manager dial plan model requires that all Cisco CallManagers located within a cluster be connected through an intercluster trunk with a route pattern for each of the other clusters within the domain.
- ☑ The Gatekeeper dial plan model helps to clean up the overhead inherent in the Cisco CallManager model. This is because the Cisco CallManager only needs to maintain one intercluster trunk, known as the *anonymous device*. This device is like a point-to-multipoint connection in frame relay, as the Cisco CallManagers don't need to be fully meshed. In this setup, the gatekeeper is able to use the anonymous device to route calls through the network to the correct Cisco CallManager (or cluster).

- ☑ The Hybrid model allows for the automatic overflow to the PSTN of calls destined for the WAN which are unable to allocate sufficient resources. It only needs one anonymous device for each Cisco CallManager (cluster), thus minimizing the overhead of having to mesh the Cisco CallManagers. It does require two routes for each destination, however, one to the WAN and one to the PSTN. The drawback is you need to configure the dial plan on the gatekeeper and the Cisco CallManager.
- ☑ For every gatekeeper located within your domain, you must configure the intercluster CODEC you would like to use, as well as enable the anonymous device. When that is complete, you will need to configure the router pattern to allow calls between clusters. You would do this by selecting a CODEC for all intercluster calls, defining the region that the gatekeeper and cluster are located in, and select the appropriate compression rate.
- ☑ When configuring the Cisco CallManager gatekeeper, you are required to enter a zone. Each Cisco CallManager will register with that zone, its zone prefix (the directory number ranges), the bandwidth allowed for each call admission, and the technology prefix for voice-enabled devices. Cisco CallManager will need the gatekeeper to explicitly specify the IP address of the Cisco CallManager within a single zone, then you must disable the registration of all other IP address ranges so it can only exist within that zone.

Video Dial Plan Architecture

- ☑ Corporate video conferencing was first introduced in the 1980s as a way to help people in different cities communicate more effectively. These first-generation solutions were based on the ITU H.320 standards defining ISDN connection-based videoconferencing.
- ☑ The Cisco Multimedia Conference Manager (Cisco MCM) is a specialized Cisco IOS software image that lets network administrators support H.323 applications on their networks without compromising mission-critical traffic from other applications. The Cisco MCM serves two main functions: it acts as a gatekeeper, and as a proxy.

- ☑ A gateway is an optional element that can be implemented within the H.323 deployment. It is an endpoint on the LAN that can provide real-time, two-way communication between H.323 terminals or other gateways. It is also capable of using the LAN and other ITU terminals located on the WAN by using H.425 and Q.931 protocols.
- ☑ A proxy gateway is a secured connection between H.323 sessions. The Cisco Multimedia Conference Manager contains a proxy as part of its infrastructure so it can provide QoS, traffic shaping, and security and policy management for H.323 traffic across any secured connection.
- ☑ The H.323 gatekeeper is an optional component capable of providing call control services to H.323 endpoints. You may implement multiple gatekeepers within your network, and they will remain logically separate from the endpoints. There are currently no standards for gatekeeper-to-gatekeeper communications, so you may want to explore other options before installing multiple gatekeepers within the same segment. You could install terminals, MCUs, gateways, or other non-H.323 LAN devices since these may coexist in the same environment.
- ☑ An MCU is a device that aids in getting calls to three or more endpoints in conference type deployments. It is usually a centralized device that assists in the facilitation of conference sessions for data, video, and/or audio.
- ☑ Video dial peers is a feature supported only on the MC3810 Multi-Service Concentrator.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: I am having problems with my dial peers, what can I do to troubleshoot?

A: You can troubleshoot your dial peer configurations with the following steps:

1. PING the IP address associated with the problem area. If you cannot PING the address, your system is either misconfigured or you have lost connectivity.
2. If this is not the problem, use the **show dial-peer voice** command to verify the system is operational.
3. If that doesn't help, use the **show dial-plan number** command to see if the information is configured correctly on the local and remote routers.
4. (Optional) If you have set up number expansion, use the **show num-ex** command and check that the partial numbers are mapped correctly to the full E.164 telephone number on the remote router.
5. (Optional) If you are using translation tables, employ the test translation-rule to verify that digit manipulation is taking place.

Q: How do I verify that I have configured the video dial peers correctly?

A: By using the command **show dial-peer video** at the exec level, you will receive information that may look something like this:

```
Video Dial-Peer 1
       type = videocodec, destination-pattern = 123
       port signal = 1/0, port media = Serial1
       nsap = 47.0091810000000050E201B101.00D07B08B6A2.C8
Video Dial-Peer 2
       type = videoatm,   destination-pattern = 789
```

```
       session-target = ATM0 pvc 70/70
Video Dial-Peer 3
       type = videoatm,    destination-pattern = 456
       session-target = ATM0 svc nsap
47.0091810000000050E201B101.00D01E82ACB2.C8
```

Please note that the second dial peer has used a PVC with the VPI/VCI pair, and the third dial peer uses an SVC. If this doesn't solve your problem, follow these steps:

1. Check the LEDs to be sure connections have been properly made, and the device is functioning properly.
2. Make sure the interfaces are not in shutdown mode.
3. Check the configuration of the interfaces.
4. (Optional) If you are using SVCs, enter **show atm pvc** from the exec prompt and check the information to see if ILMI and Q.SAAL PVCs are configured for SVC communication.
5. (Optional) If you are using PVCs, make sure they are allowed to dial PVC connections and that the connections have a service class configured for Constant Bit Rate (CBR).

Continue as needed.

Chapter 10

Designing and Implementing Single Site Solutions

Solutions in this chapter:

- **Using AVVID Applications in IP Telephony Single Site Solutions**
- **Using AVVID Applications in Single Site Solutions**
- **Using AVVID Applications in Video Single Site Solutions**

- ☑ **Summary**
- ☑ **Solutions Fast Track**
- ☑ **Frequently Asked Questions**

Introduction

Voice over IP (VoIP) solutions promise all manner of cost savings, increased flexibility over traditional Private Branch Exchanges (PBX), and faster time to production for additional telephony solutions. The major point is a faster return on investment for the solution itself. This chapter focuses upon the many factors present in bringing the benefits of Voice over IP to a single site solution. Further, it discusses the cost and engineering issues that surround single site solutions, because these issues will be the definitive mitigating factors that determine if Voice over IP is right for you.

This chapter will introduce you to Cisco's VoIP solution for single sites, how to design for it, what to watch for in the way of the network infrastructure changes, and some plain old common sense factors that might just make some decisions for you. We'll present the core Cisco solutions for single sites, but these solutions are more appropriately geared towards smaller sites in the less than 25 users category. The solutions, however, can also scale upwards to handle nearly 100 users with a few changes, which will be mentioned as the solutions are presented.

In later sections of the chapter, you'll be introduced to two newer parts of the AVVID solution—IP/TV and IP/VC solutions. These bring the benefits of broadcast television and two-way video conferencing to small and single sites, which offer the immense benefit of dramatically reducing travel costs and speeding changes out to the field users. You'll learn how these two solutions can benefit you, and how sometimes they are more of an evil than a benefit.

Using AVVID Applications in IP Telephony Single Site Solutions

In the world of corporate networks, nothing impacts the bottom line more than the telephone. It is an instrument used by every employee every day of the week, for one-on-one calls as well as conference calls. The cost of using this simple instrument varies with the abilities of the phone, with the local and long distance phone providers, and lastly, with the actual telephone and PBX to which the telephones connect.

It's no surprise that corporations are always looking for ways to save money in their quest to improve stock value, but it's really interesting that the phone system is the last to get consideration as a place to save money. Corporations always want a return on their investment, both in the design proposal and in the actual usage of these new technologies—and VoIP is no exception.

This chapter will focus upon the most critical part of a large VoIP solution—how to design, implement, and work with the single site solutions. This criticality stems from the fact that there are more sites of less than 25 users than there are large enterprises where resources are not quite as difficult to obtain. Small sites usually have only one Ethernet network, perhaps a simple router for internetwork connectivity, or maybe a small network server, with very limited IT resources.

It is for those limited resources that VoIP must have the following positive influences in order for it to be a cost-effective solution:

- Require only small network upgrades instead of a full network replacement.
- Be capable of supporting similar, if not increased, telephone capabilities, like that which the current telephone system uses.
- Require minimal administrative overhead to manage and maintain the VoIP system.
- Be flexible for new VoIP and multimedia solutions.

This isn't to say that large corporations have lots of money to burn, but rather that an investment into VoIP solutions represents a long-term dedicated effort, one that must provide significant value. This chapter will show how this level of investment does indeed have that kind of payback, ease of maintenance, as well as the freedom to accommodate change at a pace that would overwhelm traditional PBX systems.

As you read this chapter, keep in mind that organizations desiring to use VoIP and multimedia solutions must be willing to accept certain facts regarding the solution, the most important of which is having a single vendor solution. The hardware devices used to support the specialized software applications are very specialized themselves because of the job they must do. Therefore, to create the most cohesive and practical solution, the users of this new VoIP system will be using all Cisco devices.

Before deciding that single vendor solutions are bad, read this chapter and find out why the Cisco solutions presented here are not within the realm of those single vendor issues. Customers have time and again used these solutions without fear of being stranded by such specialized solutions. Instead, users and business owners alike will benefit from reduced telephony charges, the decreased time required to perform moves/adds/changes in the telephones, as well as a new flexibility for multimedia solutions that heretofore was unavailable. So, sit back and enjoy the technology review and learn how Cisco VoIP solutions work as

one part of the Cisco AVVID program. At the end of the chapter, you'll learn how to use this new AVVID system for multimedia presentations.

Designing the Voice over IP Network

VoIP solutions use the same data network as regular data transmissions, such as file/print services and e-mail. The difference comes into focus when voice packets begin flowing across this same network. This section will delve into the issues of network architectural changes that may be required to support VoIP and multimedia solutions. These issues, whether small or large, apply equally well to site designs of different sizes. Small sites, however, aren't nearly as sensitive to traffic volumes as large sites. Nonetheless, this section will present the same technical focus you'll need to apply to any network that must support voice-quality data.

Considerations for the LAN

Cisco's AVVID solutions all impact the data network because heavier data flow is now possible when all of the AVVID functions are used. To avoid congestion and assure proper voice quality, Cisco created what is known as Voice VLAN ID (VVID) that segregates traditional data from voice data. In voice-capable devices, the VVID is used to provide that segregation of data flows to assure the Quality of Service (QoS) and Class of Service (CoS) needed. Both the Ethernet switches and gateway must support these services, which means that the enterprise version of these devices will be used. Consider the traditional data network, which is shown in Figure 10.1.

This is a very simplified view of a common network, which contains users, servers, a network-attached server, and a connection to the public Internet. More advanced networks simply contain more of each of these components, and are attached together in various configurations. For this discussion, we'll evolve Figure 10.1 into a more complex network that is typical of small sites that may decide to use VoIP solutions. Next, let's consider the typical telephone system found in most small offices. Whether the office has 3 users or 30, the basic format of the office communications is the same (see Figure 10.2).

These two systems are separate and distinct in every way, and are serviced by two different service and professional providers. For the most part, the data network professionals have very little, if any, idea of how the telephone system works. The reverse is true for the telephone system maintainers. Thus, the business owner of this small site has two maintainers to work with, to order services from, and to maintain contracts with for these services.

Figure 10.1 A Typical Small Site Traditional Data Network

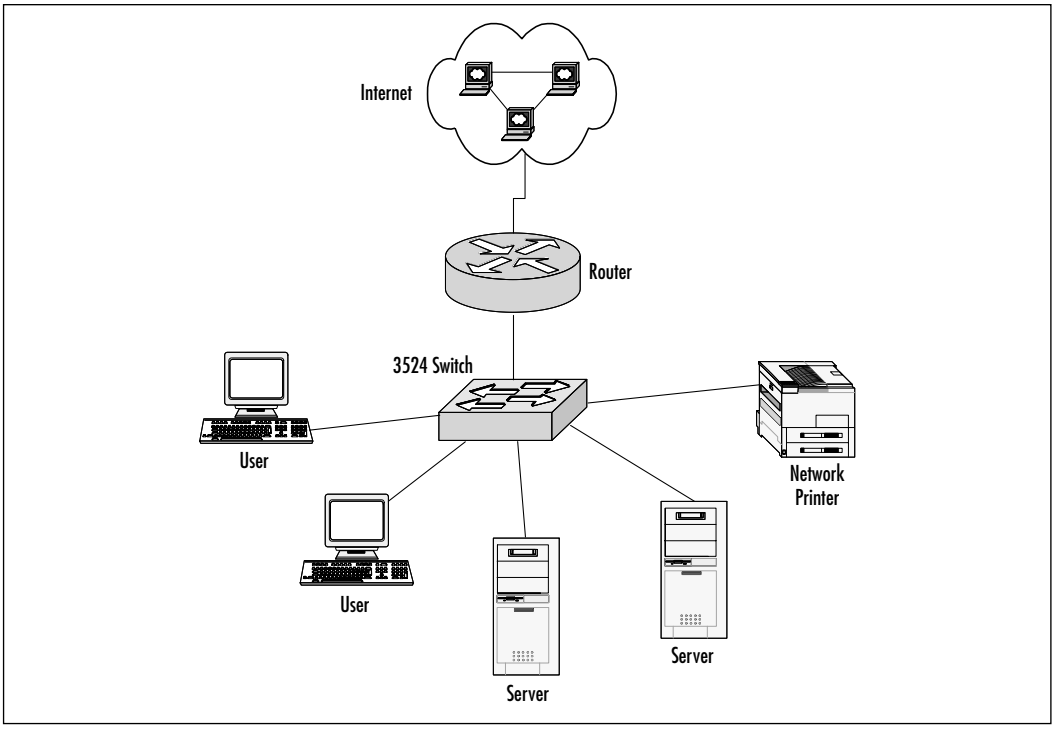
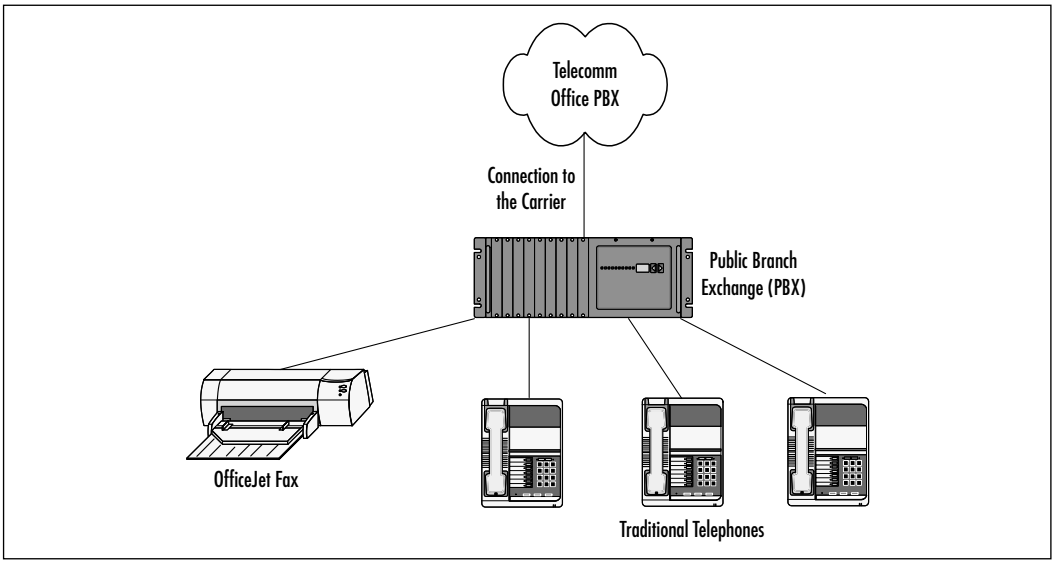


Figure 10.2 A Typical Small Site Telephone Network



Therefore, in order for this typical small site system to accommodate VoIP and AVVID solutions, these two systems must first be merged onto a suitable network infrastructure. Today's typical data network uses category-5 twisted pair cabling as a minimum, which easily supports voice and data networking. While this is a good start, the LAN must also support moves, adds, and changes for Ethernet connectivity on a very easy-to-maintain network infrastructure. This means that for a period of time, the two systems will remain apart and operate as parallel systems just like they did before all of this first started.

As the merger of the two systems begins, there is a distinct order to the migration, as stated in the following:

1. Connect the new AVVID-capable system to the external telco provider.
2. Install and configure supporting VoIP and AVVID systems onto the updated network.
3. Begin replacing the standard analog telephone devices with the new VoIP devices, usually one at a time to ensure a smooth transition.

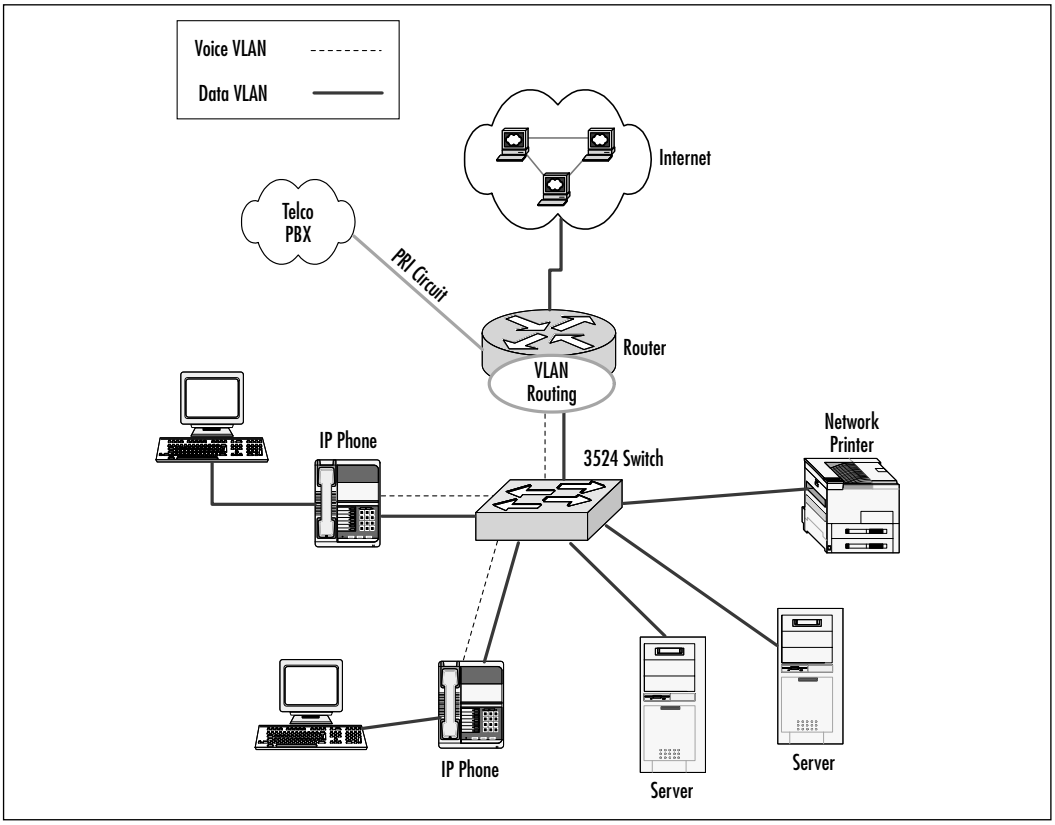
Though these seem like three short steps, it might take two weeks to complete them for 30 users. One of the most important aspects of small site VoIP solutions is to make certain the proper hardware is used the first time around so this critical capital expense is done only once. Having the proper LAN hardware also ensures that the installation and migration goes as smoothly as possible. When this migration completes, the new VoIP system will look something like Figure 10.3.

The forthcoming sections will help you understand how to perform this migration, and how to create your very own new VoIP-capable network. For now, you must understand that in Figure 10.3, the solid lines are for the data VLAN, the dashed lines for the Voice VLAN, and also realize that the router performs routing between the VLANs when necessary. This configuration makes certain that data packets do not interfere with voice packets, and ensures the proper quality of service in the networking devices required to maintain the proper voice quality. This small site solution was accomplished by deploying the following Cisco equipment:

- A Model 2621 router with 16MB of flash memory, Cisco IOS version 12.1(5)T8, 48MB of memory for the operating system and shared buffers. The IOS you use will probably differ from this due to your own requirements.
- One Primary Rate Interface (PRI) module for the telco central office connection.

- One Model 3524 In-line power Ethernet switch to provide power for the Cisco Model 7960 IP phones.
- Cisco Model 7960 IP phones, which support multiple lines, speed dialing, conferencing, and multiple feature support. This phone is actually a two-port Ethernet switch that provides 10/100 Mbps Ethernet connectivity for the desktop computer, as well as connectivity to the 3524 In-line power switch.
- A Cisco CallManager Server, which provides the core PBX functionality.
- A Cisco Unity Messaging Server, which provides voice mail capabilities interconnected via Microsoft Exchange Server v5.5 for voice/messaging interaction.

Figure 10.3 The Newly Merged VoIP Network



Connecting the Site to External Telephony Systems

Thus far, we've talked about the LAN environment and its structure when used to support VoIP and AVVID solutions as a whole. To make this a complete solution, the LAN VoIP users must be able to connect to the outside world. In traditional PBX systems, the usage of a PRI provides up to 23 channels of 64 Kbps across a standard four-wire communications circuit. This PRI, depending upon the location in the country, can cost from $20 to $60 per single 64 Kbps channel. Even at its cheapest, such connectivity adds a monthly recurring cost of $400 to the solution, not counting long distance charges. However, you might find pricing for a PRI different from these figures, which are well established in the southeastern United States. These figures are provided merely as a reference point.

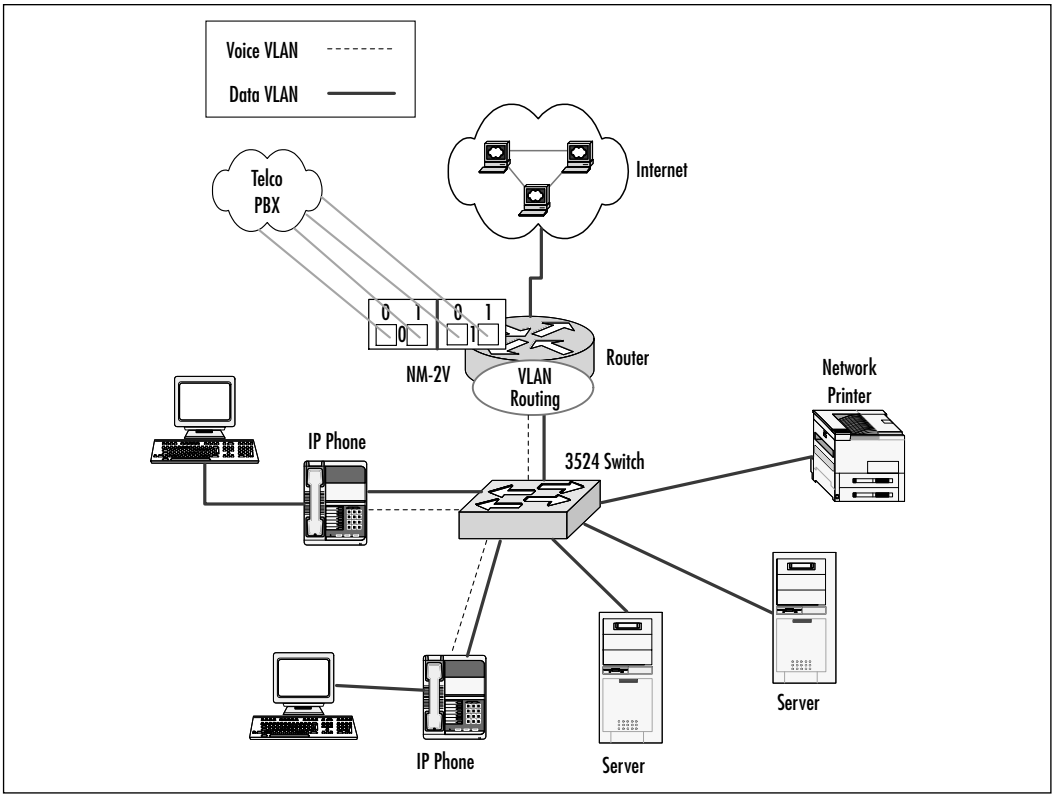
Since the site will pay long distance charges regardless of the solution, whether it be VoIP or a traditional PBX system, this charge can be ignored when considering the Return On Investment (ROI). There is, however, a cheaper solution for small sites without having to use a PRI circuit. Most small Cisco routers support the usage of the network module voice slot in either the one- or two-slot design, the NM-1V and NM-2V. Each slot can accommodate one of the following three modules:

- **FXS module** Foreign Exchange Station (FXS) is used to connect to end station devices such as analog telephones and analog fax machines.
- **FXO module** Foreign Exchange Office (FXO) is used to connect the VoIP network to the outside world via standard analog telephone lines, which are much cheaper than a PRI circuit.
- **E&M module** Ear-and-mouth (E&M) is used to provide trunk connections between VoIP systems.

Bypassing the cost of the PRI is one major accomplishment in realizing cost savings (use Figure 10.3 as a reference). Instead of using the PRI, the NM slot would use up to two of the voice modules to provide a total of four connections as shown in Figure 10.4.

Now that standard analog phone lines are in use, we've reduced the cost of the VoIP solution for this small site dramatically, but at a cost. This modification to the solution means that no more than four active conversations may occur at any given time, whether or not these are inbound, outbound, or a mix of call types such as voice and fax. This limitation is a trade-off for the cost of the PRI circuit.

Figure 10.4 The Updated VoIP Solution Using FXO Modules



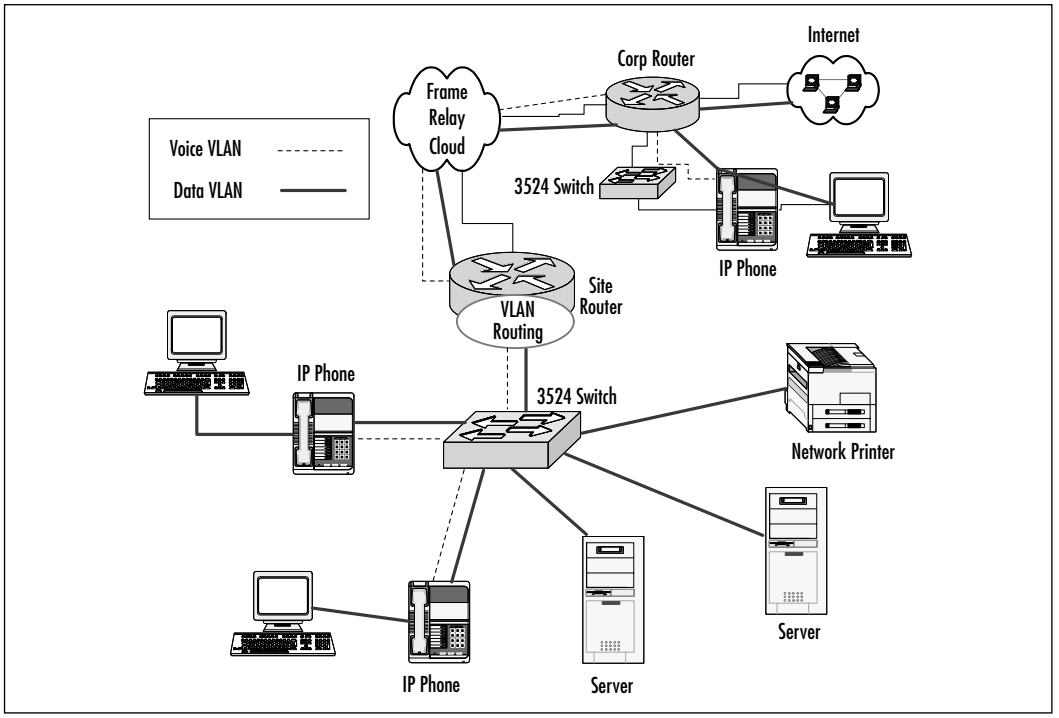
Connecting the Single Site Back to the Corporate System

Instead of using either a PRI or the aforementioned network module slot for the analog lines for connection to public telephone services, some sites connect back to their head office, or its main network by way of a frame relay network. This type of circuit provides dedicated connectivity for the small site to gain access to the head office's resources, such as a mail server, or for consolidated access to the public Internet.

This type of connection provides a number of benefits, including more stability and independence over the telco providers, flexibility over the routing of data and voice, and reduced cost. Even with the cost of the frame relay circuit, the majority of corporate phone calls are inter-office and could use the frame relay circuit between the small site and the corporate network. But, connections like this mean that the small office must take its Internet connection from the

head office unless the small site has its own cost-effective Internet solution, like that shown in Figure 10.5.

Figure 10.5 The Small Site Taking Its Services from the Head Office



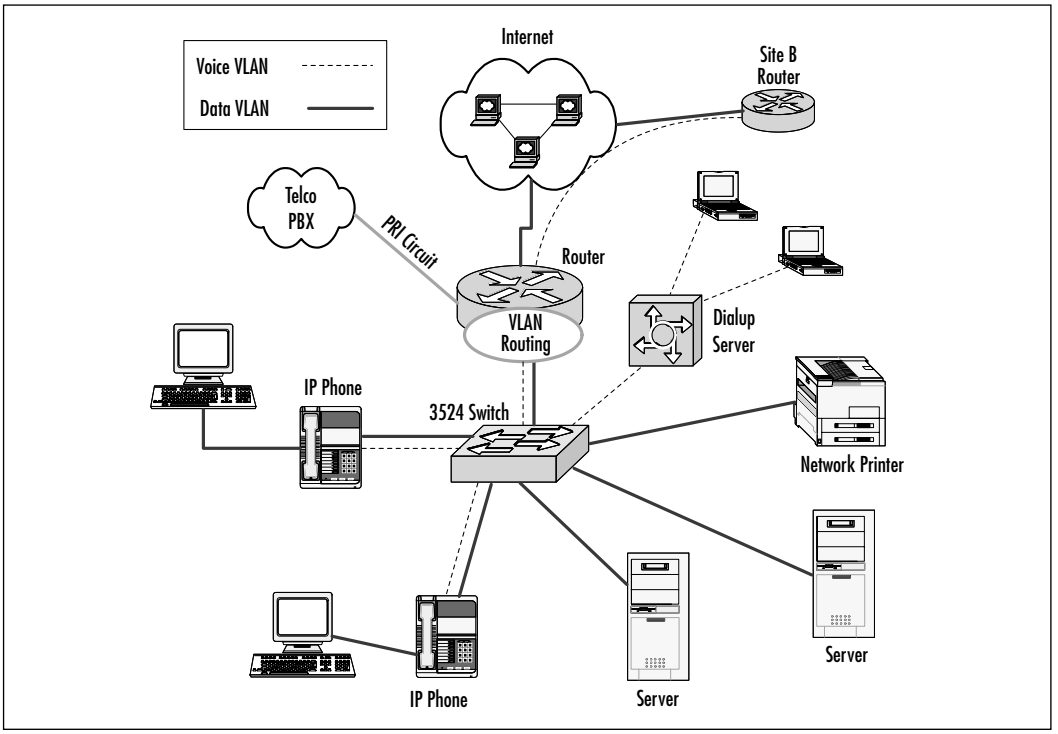
For the sake of the inherent security of frame relay circuits, this type of arrangement has the added benefit of using as much of the circuit capacity as is needed, up to the connection limits purchased from the provider. This configuration also means that instead of placing a CallManager at the small site as well as on the head office network, only one CallManager is required at the head office network to serve both the head office and site IP phone services. Before you decide on such a solution, you must ensure the frame relay connection and all interconnecting devices are rock-steady and have a stable configuration.

Connecting the Single Site Back to Other Small Sites

There are times when connecting a small site back to the head office is not possible, much less financially feasible. Some sites are closer to other small sites, and can serve as a stepping stone to the corporate platforms. Before choosing this type of configuration, you must ensure that both of the small sites have sufficient

resources to handle the call volume. This configuration is not only possible, but can bring significant cost savings if the two sites are very close together yet cannot be located in the same building. This is useful where many mobile users reside, yet no single office exists. The best use for this is where mobile users dial in to one site, and use the Cisco IP SoftPhone on their computer or laptop, as shown in Figure 10.6.

Figure 10.6 You Can Provide IP Phone Services via a Dialup Connection



However, notice in Figure 10.6 the dashed line between the two site routers. This is an IP Security (IPSec) site-to-site Virtual Private Network (VPN), such that each set of devices on each network appears to just be another device on a larger network. These devices are able to communicate together, use the same resources, and place IP phone calls between the two sites. One user on one network would have no idea that the data is carried between the two sites by way of secured communications across the public Internet, but that's exactly what is happening here.

An inescapable issue with this type of arrangement is the possible loss of connectivity between the two sites should any manner of problem arise with either

site's router, connection to the public Internet, or the VPN itself. Worse yet, many small sites do not have a properly sized Internet circuit capable of carrying not only the VPN traffic but traffic destined for the public Internet. Sites that are on a VPN connection back to their head office typically use at least a 256 Kbps or faster circuit, but may be limited to as low as 64 Kbps, such as ISDN.

When these lower-speed capacities are present, connecting two sites together for the purpose of VoIP or other AVVID solutions becomes very challenging. If these slower connection speeds cannot be increased, then running AVVID solutions between the sites will not be possible. In these days of x Digital Subscriber Line (xDSL), even the slowest IDSL speed of 144 Kbps is capable of supporting just the VoIP portion of the AVVID portfolio. You can also bond a pair of v.90 modem dialups into a 112 Kbps channel between a pair of Cisco 2600 class routers that use asynchronous modems.

Choosing a Voice-Capable Gateway

Now that you understand some of the pitfalls and pleasures of using VoIP solutions with small sites, you need to choose the proper gateway router that controls the VoIP system. This section will discuss a few of the VoIP gateways available. While there are many other available gateways, these solutions will revolve around the small site solution.

Types of Voice-Capable Gateways

A clear definition of a *voice-capable* gateway is a router that provides not only data services, but runs the proper Cisco IOS firmware that provides voice services. These services are, in their basic form, the following topics:

- Controlling and utilizing Digital Signal Processors (DSP) for processing analog calls
- Providing call processing to a CallManager
- Providing routing for VLANs between voice and data subnets

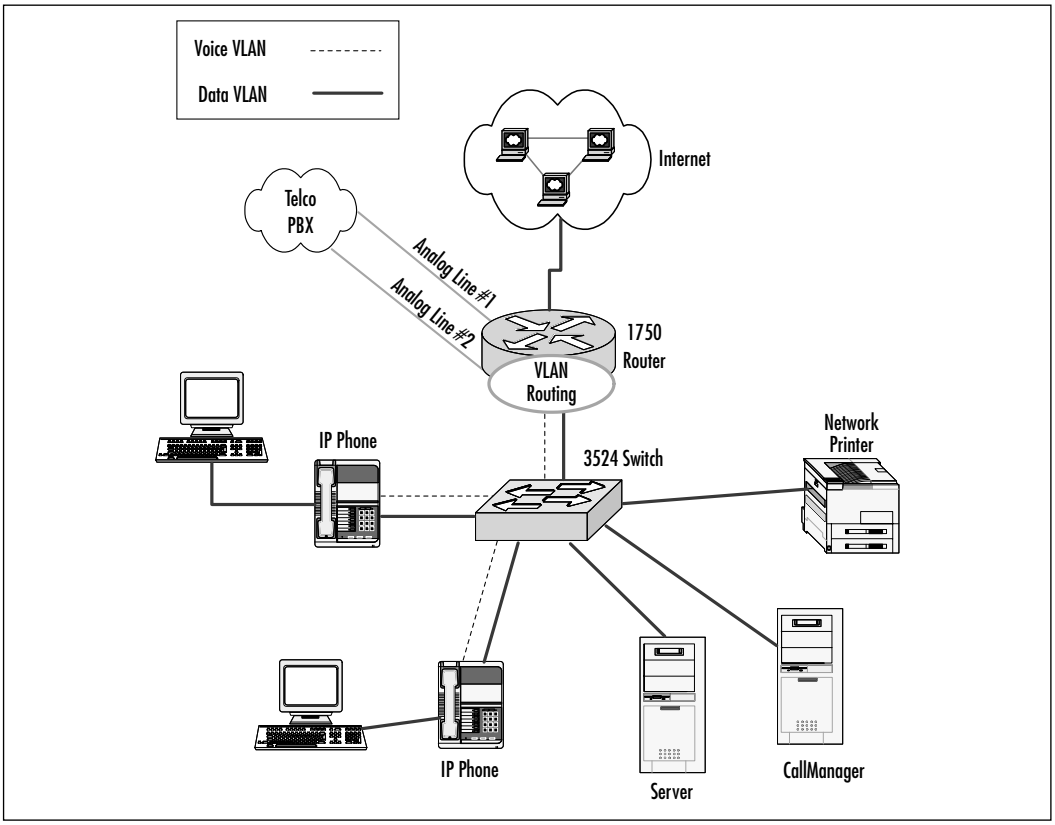
A voice-capable gateway must have sufficient CPU processing power and system memory to handle these functions, as well as any other AVVID services that may arise. This is where the majority of problems occur in new VoIP designs because the wrong gateway is selected. In some solutions, sites will try to use the same voice-capable gateway for both voice and data services. This means that the same router provides the telco connection, Internet access point, and VLAN routing in the same gateway.

The best solution in this environment is to use a Cisco 1600 Series router for the Internet connection, and then deploy the 1750 router with the voice-capable IOS to handle the VoIP solution. This keeps the system routing clean and distinctly separated from data services. If the site cannot afford to have two gateways in this type of arrangement, the 2600 Series can perform both data and voice services provided that it has sufficient memory and that the proper voice hardware is installed.

Cost-Effective Gateways for Small Sites

When small sites do not have the financial services needed for more expensive devices, there are several Cisco solutions that will provide the bare essential VoIP solutions. The most basic need is for one Ethernet connection, and one or more Plain Old Telephone System (POTS) connections engineered for at least one analog telephone line. Figure 10.7 shows this simplified site drawing of how the Cisco 1750 router can perform both data and voice services for a few users.

Figure 10.7 The Barest of Small Site Connectivity



This solution offers a much-reduced cost when it comes to gateway selection, yet provides the minimal VoIP solution. This Model 1750 router is the smallest in the Cisco line that provides full VoIP capability along with the best selection of hardware adapters for small site connectivity. By using POTS analog lines, virtually any small site can have a degree of VoIP benefits without the major costs associated with site-to-site or site-to-major backbone connections.

Cisco IOS Solutions for Voice Gateways

To select voice-capable hardware is not enough. You must also choose the correct IOS firmware for the gateway router so the gateway can speak the proper voice lingo to CallManager. Savvy network designers use all facets of the Cisco Web site to learn as much as they can about the products so they can choose the proper equipment. This portion of the Cisco Web site is called "Cisco Connections Online" or CCO for short. Access to CCO requires that you have an account with Cisco to access this private area. This account is usually granted for customers that purchase the SmartNet maintenance when they purchase their Cisco products. CCO grants you access to special areas of neat documents, technical tips, and tools for searching the feature sets of IOS versions.

To find the Cisco-approved IOS for our small site, we go to the Feature Navigator in Cisco's Web site. For our small site, we've chosen the 1750 as our voice-capable gateway. In Feature Navigator, we first had to type in the feature we wanted, which is Media Gateway Control Protocol (MGCP). When MGCP was typed in and the search began, we were presented with four optional results. One was the 1750 voice-capable gateway and the other was MGCP support for CallManager on other gateways like the VG200 and 2621 routers.

We then selected VoIP signaling for the 1750, and told the Web site to continue. What we got next was a set of dropdown menus to begin narrowing down the choices. Clicking the release drop-down, we see that there are only two possible IOS choices, both in the 12.2 family for the 1750 gateway. We chose the 12.2(2)T family, the T meaning "technology" IOS. The T code has all of the newest features such as VoIP, but requires much more memory and flash than does say the plain IP-only IOS. In the platforms dropdown, we chose our 1700 family of gateways, and lastly chose the IP/ADSL/VOICE/Plus code.

Even though we won't use the ADSL portion of the code, ADSL is included with all 1700 Series gateways. This yields the following IOS for us to order with the new gateway:

```
C1700-sv3y7-mz.12.2-2.T
```

This is the easy way to find the Cisco-approved IOS for your site. We could do much the same thing for say a 2621 router, but in Feature Navigator choose the MGCP support for CallManager. After choosing the MGCP Support for CallManager and going to the platform family, notice that the choices are much different. Among the possible voice-capable gateways are the 2600, 3600, ICS 7700, ubr905, and the VG200. The first three selections are families of gateways while the last two are individual devices. Also apparent is the slimmed down choices of IOS versions since we're looking at voice. Regardless, we've now selected the correct Cisco-approved IOS for our new voice gateway.

Problems Using the Voice Gateway for Combined Data Access

Lastly in this section, we shall discuss using one gateway for both data and voice access. We've seen many sites that, because of financial constraints, use the same gateway for both voice and data services. While the intent is good, the integration of this idea is usually marginally operative at best. The reason for this is not in pushing the gateway's CPU to maximum performance, but rather in running too many services on the same gateway. At times, the gateway can be confused and cause reboots at the most inopportune times. This isn't to say that every site that uses the same gateway for data and voice will have this problem—far from it—but you should be keenly aware that this situation might exist. Figure 10.8 shows the recommended site configuration when this issue arises, regardless of the amount of users.

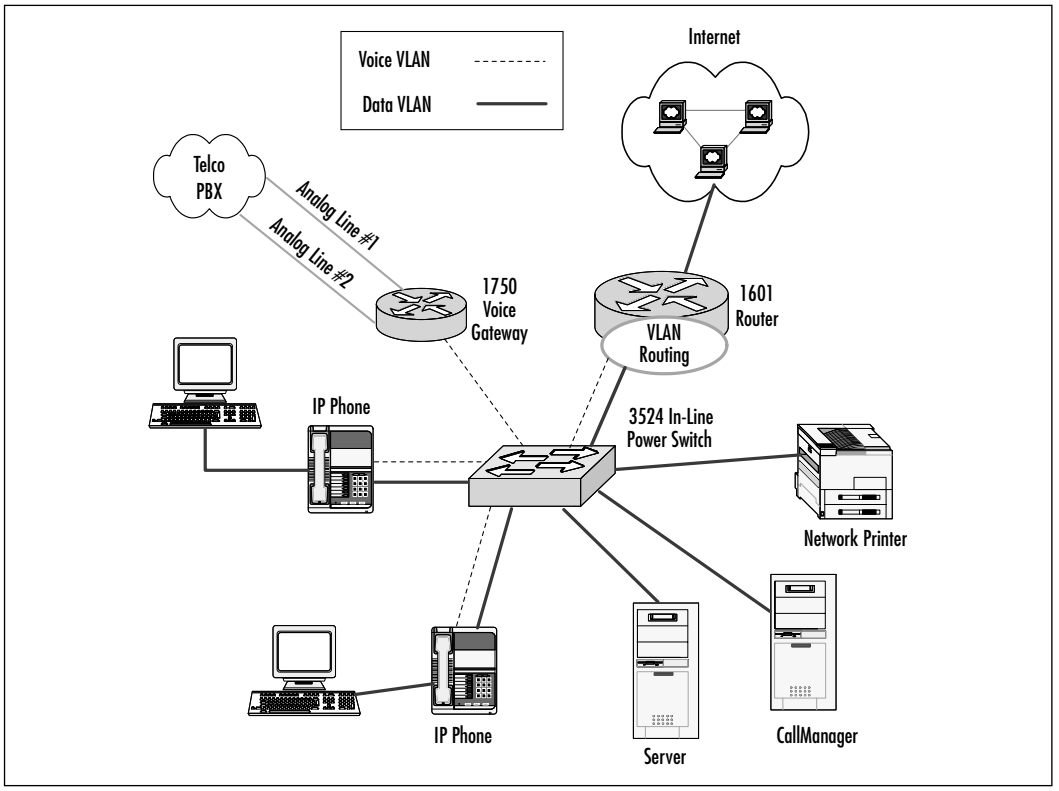
The most important idea to recognize in Figure 10.8 is that now the voice is not only separated by VLAN, but by the gateway as well. We've reduced the overall site cost for electronics by about one third while retaining the equivalent functionality. If this site were connected back to the head office, or to another site that has a CallManager, then this CallManager can be eliminated, further reducing costs. But, if this site must have dialing capability regardless of connectivity to other sites, then a CallManager at this location must exist.

Modifying an Existing Network to Support Voice over IP

The previous section discussed the issues surrounding a voice-capable network, but an even more important issue exists: adapting an existing network infrastructure to support VoIP solutions under the Cisco AVVID umbrella. This section will extend the previous ideas into political decisions regarding how to modify an

existing network so it supports VoIP should the decision be made to go forward. It must be understood that once this decision is made, most actions are neither reversible nor stoppable without a loss of capital investment. Perhaps the only action outside this rule would be an upgrade to the wiring infrastructure.

Figure 10.8 The Recommended Small Site Gateway Design



This Must Be a Pure Cisco Solution!

We've heard more than once the comment concerning integrating a solution using one and only one vendor, and how this locks them into that vendor's devices. This is true of any solution, including the Nortel Networks VoIP system. Therefore, it can be safely said that the Cisco solution is no more risky than the Nortel Solution—they both provide VoIP capabilities.

To reap the benefits of the VoIP solution, you must run end to end Cisco devices with compatible IOS and Catalyst firmware that supports VoIP and MGCP. This includes the switching solution for the VLANs, and the IP phones as

well. The IP phones are powered by three different methods with varying costs and trade-offs:

- **Inline powered switches** These are switches like the 3524 In-line power model. You must specifically request the in-line power model when ordering these units. This device removes the need for external power adapters and is compatible with a PC's Ethernet adapter. It costs much more than other solutions, but is the cleanest and has the least maintenance of any power solution.
- **Separate external power converters** Used for IP phones, they are similar to those employed for laptop computers. This is the cheapest power solution, but uses the cumbersome external power converter. If you go this route, you should keep several of these around for spares. This external power supply is sometimes known as a *power cube* device.
- **Powered patch panels** These are special category-5 patch panels that provide power to the jacks of the patch cords much like the 3524 In-line power switch does, except through the rack mounted patch panel. These often fall between the cost of the switch and the external power supplies, and are fixed in capacity and size.

Lastly, you must use the Cisco CallManager solution, which is a customized Compaq server running Windows 2000. The version of Windows 2000 on this server is also specialized for the Compaq server, and cannot run services other than the required DNS, Trivial FTP, and DHCP functionality required for the phones. This seems like quite a restraint, but it ensures that the CallManager server is not burdened with unnecessary processing requirements given that CallManager might have to service as many as 2,500 simultaneous call processing connections at a time.

Let's talk briefly about the various IP phones Cisco produces. There are six types of phones, which can be broken down into two functional groups. The first group is the older first generation of IP phones such as the Model 12 and VIP Model 30 programmable phones that had multiple lines and as many as 30 memory phone number settings. The second generation of IP phone is the 7900 Series phones, the 7910, the 7910+SW which has a pair of 10/100 switch ports, the 7940, and 7960. The 7910 is analogous to the single line unit with few memory positions and only one line. The 7940 supports two lines and an increased number of memory storage positions while the 7960 is considered the executive phone. Having as many as six lines, the 7960 supports the XML standard to enable special

features and functionality on the LED screen, such as lightweight messaging. The 7940 phones also support the XML standard.

Regardless of the model you choose, these are all Cisco proprietary phones that work well on a Cisco-powered network for VoIP solutions. Prices ranged from $150 to $600 per handset in July of 2001, but varied more widely when items were purchased in volume.

Deciding Which Type of Public Telephony Access to Use

One very important piece of the puzzle to consider is the type and size of the external telco connection to use. This is mandatory since the internal users must reach the outside world. The two accepted types of connectivity are permanent leased line and dialup connectivity. This is further broken down into the following functionality types:

- **T-1 Primary Rate Interface (PRI)** With a total of 1.536 Mbps capacity, this link is broken down into 23 channels of 64 Kbps capacity. Each phone conversation will utilize one channel regardless of how much of the 64 Kbps is actually used. Therefore, one T-1 PRI can host 23 conversations simultaneously. This type of connection costs between $28 and $60 per channel depending upon the provider you use, given that you're getting business class of service and circuit stability.
- **ISDN Dialup** Consisting of two 64 Kbps channels, this dialup technology can handle up to two calls at one time, each using 64 Kbps of bandwidth. Be careful when choosing ISDN. Costs can vary wildly; from between $40 a month fixed rate to as high as $700 a month when all the surcharges are added up.
- **POTS analog lines** These are the standard phone lines found everywhere. Be careful you don't get hit with business class rates, however, which can be as high as a PRI channel without the stability of a PRI circuit itself. Standard POTS lines usually average about $18 a month for basic service, whereas the same line used for business services can easily exceed $45 a month plus per-minute usage fees.
- **xDSL circuits** Becoming more popular than ever, more data and voice companies are offering Voice over xDSL, which is nothing more than using a portion of the circuit to transmit and receive voice traffic between the end user and telco's PBX system. The Cisco 1750 gateway

is one such device now supporting ADSL for voice, as is the 2600 class of gateway.

Deciding which technology to use is a matter of having previously decided upon a voice-capable gateway, getting quotes from your local voice carriers, and obtaining the cost of installing the appropriate hardware in your gateway. Our local site decided to use the 2621 gateway with IOS v12.1(8)T, the NM-2V network module which hosts our two VIC-FXO cards providing a total of four POTS analog lines to the telco PBX. In this manner, even if our office had to move, the same gateway can be used at any new site location independent of the availability of PRI or xDSL circuits. The only drawback was we couldn't support more than four active calls simultaneously, regardless whether the calls were inbound- or outbound-initiated.

Performing a Network Assessment of the Infrastructure

Having read the previous sections, it should be crystal clear that performing an assessment of the network infrastructure is not only vital to the success of the VoIP installation, but also to the continued success of the installation. Using the word "assessment" often conjures up impressions of tens of thousands of dollars in consultant expenses, but it doesn't have to be. The tasks carried out most often in an assessment of this degree are as follows:

1. Test and validate the network wiring to assure it's at least category-5-compliant, without faults or errors.
2. Review and document the existing network electronics to determine what make, model, and part number of the device is installed. This will be used to determine the lateral Cisco replacement part, as a minimum, although a somewhat higher level of functionality is usually required.
3. Review and document the current telephony solution to determine exactly which portions of PBX is to be replaced, augmented, or supplemented with VoIP services. The current user's dial plan should be clearly documented and understood so that the correct VoIP dialing architecture can be designed.
4. Lastly, determine exactly which telephony services are required, such as voice and fax. You'll specifically need this since, from a technical standpoint, a fax call is merely another form of using one 64 Kbps channel of communications. If you have to dedicate a channel for a fax, you're

better off using a dedicated fax line instead of one of the channels previously mentioned.

From this simple list, you should now understand why an assessment is needed at one site and not needed at other locations, while still yet other sites may need only a partial site review. A customer's needs will vary between desires of using VoIP and/or AVVID solutions, but the review will just make sure that no one is caught unaware of special circumstances that may cloud the overall design.

Engineering a Mixed Vendor Solution

Given the previous discussion, it should now be clear that a mixed vendor solution should be approached with caution. There are circumstances where a mixed environment may actually work, such as installing VoIP at a small branch office where a single IP subnet is to be used for up to 20 users. In this case, any compliant Fast Ethernet switch would work fine without the need for establishing VLANs, except that you'd need to use the external power adapters for the IP phones themselves. Another issue with mixed vendor solutions is that even on a flat network, a third-party Fast Ethernet switch might not be configurable for port-based Quality of Service (QoS) or for Type of Service (ToS) tagging that VoIP solutions sometimes require for proper operations. There are some solutions where this is not an issue, and yet others where it ends up as a complete catastrophe.

The main point to be made is that if you decide on a mixed vendor solution, you inherently accept the risks of having something go wrong with the installation. When this occurs, Cisco isn't likely to be of much help to you troubleshooting other vendor's equipment and problems. Cisco isn't being rude about it, just realistic—they have no idea how well that vendor's solution might or might not be.

Using AVVID Applications in Single Site Solutions

Now that the hardware solutions are defined and out of the way, this major section will be devoted to discussing the application side of the VoIP solution, which includes CallManager, Unity Messaging, and the usage of the Cisco IP SoftPhone solution for mobile computers as well as users who do not need a desktop phone.

Using Cisco CallManager

At the heart of any telecommunications system is a device responsible for performing call management: the Private Branch Exchange (PBX). The PBX system is nothing more than hardware with an operating system that recognizes when someone starts to place a call, determines what number the person is dialing, and then determines what piece of hardware in the PBX system it will use to route the call to the destination. If the call is local, the PBX operating system understands that the destination is local, and does not route the call to outside resources.

This determination is based upon what is known as the dial plan, and the Cisco CallManager is the software component of the VoIP system that makes that determination. In another designation, CallManager is sometimes knows as an IP PBX system. The dial plan is merely the configuration of the CallManager such that the site's area code and prefix is used to help CallManager determine if the destination call is local or outside of CallManager's control. This section will discuss CallManager, its features, and how it works to control calling behavior.

Understanding the Component Parts of CallManager

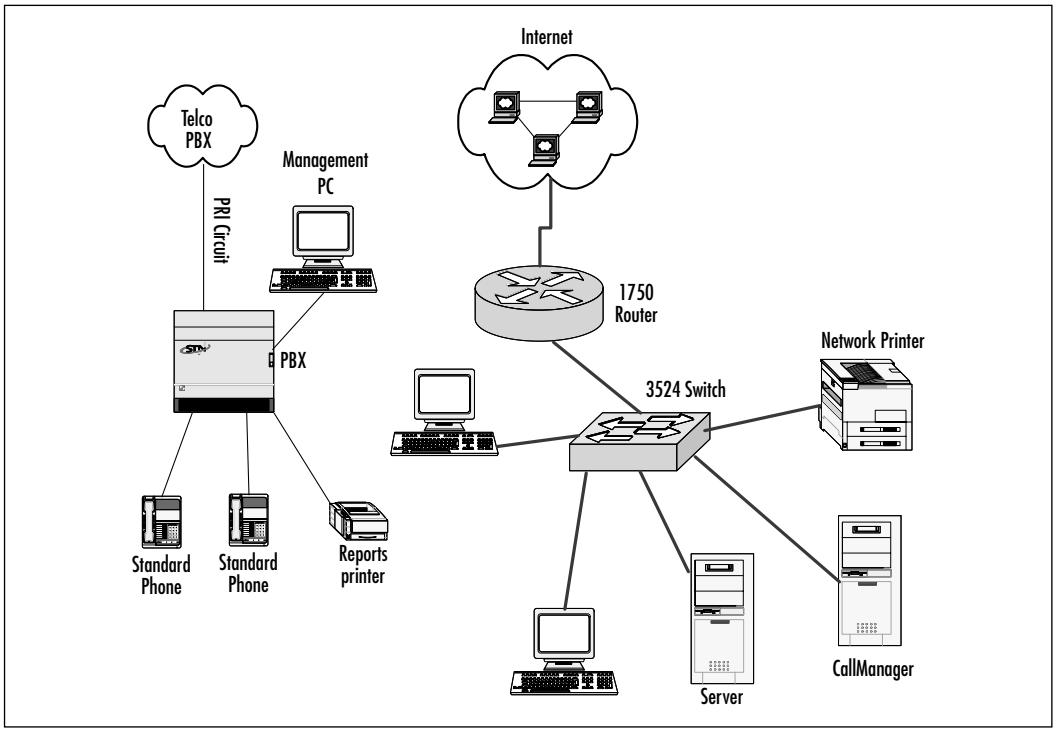
CallManager is broken down into several functional elements:

- **System controls** These areas are used to configure, manage, and troubleshoot CallManager as well as its underlying server tools.
- **Basic networking functions** While not exactly CallManager functions, CallManager runs on top of Windows 2000 Server with only the most basic networking functions. CallManager requires DNS server services, TCP/IP networking, Windows Networking, and nothing more. If the server (a Compaq in this case) requires special drivers or services, then these are in service.
- **Device controls** These functions are used to create, control, manage, and organize the IP phones into logical groups and call routing. Among these controls are call regions, device pools, and location controls for determining the type of call digitization and compression.
- **Gatekeeper and gateway controls** Used to define and control the acceptance and routing of calls.
- **User management** Creates, manages, and controls users in CallManager.

Within CallManager, all functions can be lumped under these five major areas in one way or the other. To the extent that you can draw a parallel to the standard

PBX system, Figure 10.9 shows how traditional telephony systems work for small sites and provide much the same controls as the traditional PBX system.

Figure 10.9 The Age-Old Traditional PBX System



The desktop phones physically reside on the same desk as the user's PC, yet are connected to a distinctly different system and access the site through other means. When you pick up the handset, you'll hear the dial tone provided by the PBX. When you type in the number you want to call, the PBX makes the decision as to how to route the call. You've previously seen where IP phones use the same network infrastructure as the data devices. This changes slightly because CallManager now replaces the traditional PBX previously shown in Figure 10.8.

Installing CallManager

Installing CallManager is the easiest task of any you'll experience. The installation CD shipped with the product is an automated script that performs every task required to create an operational CallManager. Do you remember what was said about CallManager using Compaq specialized server hardware? The CallManager

installation CD uses that specialized version of Windows 2000 that has the correct drivers for the server hardware, SCSI devices, and the Compaq motherboard support drivers.

The installation CD also contains the scripted installation for CallManager and the MS SQL Server v7 used for the database services. The server installation and CallManager installation are one and the same, being that the files are all extracted from the compressed CD data files. During the installation, the server will have to be rebooted several times as various parts of the system are installed and configured via the scripts.

Once the automated installation is completed, the new CallManager server will reboot one final time and present you with a completed system. CallManager has not been configured, which is what you'll begin doing in the next sections. You'll first complete the basic configuration for the hardware, then you'll perform the more advanced configuration for the users and phones themselves.

Performing Basic Configuration Tasks

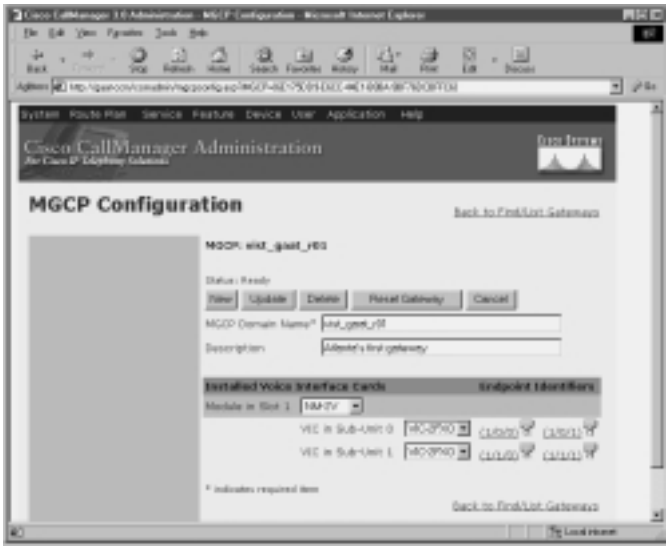
The first thing you'll need to do is log into CallManager, either locally or remotely via your Web browser. You can access it by typing in **http://localhost/ccmadmin** in your Web browser, or by replacing the localhost designator with the IP address of the CCM. Changing the administrator's password is the first action you should take, ensuring at least some level of security for the server.

For CCM to route calls, it must now know where the gateway is that will handle both call completion and CODEC actions. In the "Devices" section, you'll need to define a new gateway for CCM to use. When you do, use the host name of the gateway, not the IP address. When the gateway is added, CCM will attempt to contact the gateway and initiate an MGCP session to validate the connection. The primary means of validating the session is to check that the CCM and the gateway both belong to the same MGCP domain, which is one measure of security for CCM (see Figure 10.10). The MGCP domain is defined by the host name of the gateway, and must be the exact same name in CCM (it's case-sensitive).

WARNING

If the MGCP domain name is wrong in CCM, or if someone changes the host name of the gateway without changing the gateway's name in CCM, CCM will not be able to establish an MGCP session with the gateway in order to complete any call. Furthermore, there are no error messages in CCM to indicate that this misconfiguration has occurred.

Figure 10.10 The Gateway Configuration



If calls cannot be placed both internal and external to the site, your best and quickest troubleshooting method is using debug commands in the gateway. We'll get into these issues much deeper in the troubleshooting section.

With CCM and the gateway now communicating, we'll next need to tell CCM what hardware to use to reach off-net calls. Our earliest example uses the 2621 router with one NM-2V module, and two FXO cards to supply four ports of analog POTS lines. You'll need to configure each FXO port in the gateway to choose the order in which the lines will be used, the type of port, and if the port is to be used at all. You can activate and deactivate each line as needed should a port go bad, or perhaps is being tested by the telco office.

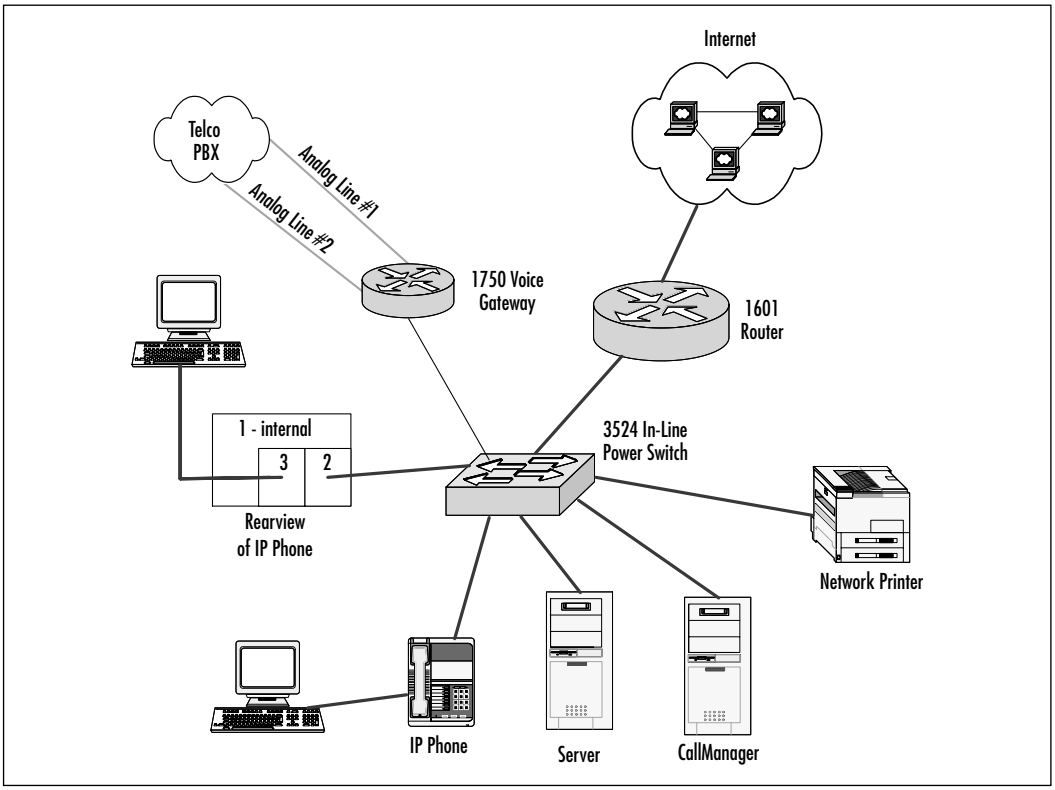
In the dial plan section of CCM, we must state how the number strings are to be treated. We selected the **PreAT** feature and used the string 9@ to reach an outside line, while denoting nothing for an inside connection. PreAT means we must dial a 9 to get an outside line (to use the gateway's MGCP services), stripping off the numbers before the @ when the call is completed. In this example, the only thing to strip off is the 9 to get an outside line, since you'd not want to transmit the 9 to the local telco as part of the dialing string. Such custom dialing strings can be used to choose a particular long distance carrier and force long distance calls to take a specific path as the preferred route.

Next, we need to make sure the DHCP server on the network functions and is properly configured to support the IP phones. DHCP is one of the few services

CCM operates, but we already have a DHCP server running on the same subnet. This allows us to provide DHCP services to the client computers and remain independent of the CCM itself. Also, all CCM servers must have an entry in the DNS server for this site. One important change to the DHCP configuration, however, is to specify the usage of a Trivial FTP (TFTP) server for the IP phones in order to download their firmware configuration. This configuration is stored on the CCM server who's IP address is used as the TFTP server in the DHCP configuration.

With both DHCP and DNS servers properly configured, we next have to physically install the phones. For the purpose of this discussion, we'll presume that the installation is on a small site without VLANs. We'll also use the Model 7960 IP phone, the most popular business class phone in use today. This model is actually a three-port Fast Ethernet switch. One port is an internal only switch port (used by the phone's electronic controls), the second is the inbound connection coming from the 3524 switch, and the third port goes to the desktop computer (as shown in Figure 10.11).

Figure 10.11 7960 IP Phone Connections



Many network infrastructures today only have one category-5 cable running to each desktop location. While there are a variety of reasons why two or more cables are never connected to each desktop, Cisco answered this issue with the release of the 7960 phone as shown in Figure 10.10. This is also why the previously mentioned network assessment is so critical to the success of any VoIP project. With only one cable installed, it must function correctly 100 percent of the time.

Even though this sample installation doesn't use VLANs, the 3524 switch and the phone tags each voice packet with the proper Type of Service (TOS) in the header of each voice packet. This ensures the switch and router properly recognize and process the packet for what it is, and don't treat it as a pure data packet. Make sure you connect the proper cables to the proper ports, each is labeled as such.

Because the phones are powered by the 3524 switch, the phone will try to initialize and boot up as soon as it is plugged into the network. The bootup operation is simple, but takes a few minutes to complete. In the first step of the process, the phone completes the physical connection to the inline power switch. The switch then sends a low voltage transmission down the wire to the phone, and the phone responds to the increased voltage by completing the return path back to the switch. The switch sees this as an acceptance of the increased voltage, and so ups the voltage again by a small increment. The phone again accepts this increase, and the process continues until the proper line voltage is present to power the phone. If this end device were a normal data device, such as a laptop, the computer's network adapter would not respond to the initial increase in voltage, informing the switch that a standard data device was now connected.

The phone attempts to get an IP address via DHCP (which has the pertinent settings) so the phone knows how to communicate across the network. One of these settings denotes where to find the TFTP server, which contains the boot file for the phone. Once the phone downloads its configuration file, which is stored on the CCM, it now knows how to contact the CCM for the rest of its configuration settings. Anytime the phone is disconnected and reconnected, the phone repeats this process. The IP phone is now registered with CCM, and will show up in the device listing whenever you add a new phone to CCM. Complete the rest of the physical phone installations, and you're ready for the next step: creating the basic dial plan and adding users.

Configuring & Implementing…

A Word about Regions and Device Pools

In the system configuration of CCM, you have to create device pools and regions that deal with how the phone call is treated. These two conditions do more for quality of service than any other configuration. There are two basic types of compression and voice handling: g.711, which uses the full 64 Kbps PRI channel when high bandwidth is available, and g.729, which compresses the voice packet down to 8 Kbps for transmission across low speed WAN links such as 56 Kbps frame relay. There are, however, several other compression types, some that go as low as 5.3 Kbps, but that require more advanced (and more expensive) DSP modules. When these high-complexity DSPs are employed, you'll need to use gateways such as the Cisco 3600 and AS5000 Series devices.

Where and when would you need such high level hardware, though? The previously mentioned 1750 and 2600 class gateways using the voice modules can provide adequate voice compression and mixing for up to four conversations, but fall short when more than four simultaneous conversations are needed. The 3640 gateway, for instance, can accept up to 12 DSP modules that each have three individual DSP processors, with each DSP processor handling one conversation. Also, when conference calls and bridging are needed, one DSP processor is required for every three participants in the call. DSP processors can be co joined for larger conferencing needs, but require the usage of more capable gateways such as the Catalyst line of switches. These Model 4000 and Model 6000 Series Catalyst switches utilize the 8-port T-1 DSP module, with each module supporting three individual DSP processors. However, these 24 combined processors provide much more VoIP capabilities than do the lower end gateways.

This diatribe is not meant to say that the 1750 and 2600 class gateways are not sufficient to do the job—far from it. They each have their particular place in life as well as an associated cost. Chapter 11 will go into these more advanced issues in greater detail.

Performing Advanced Configuration Tasks

Let's do a quick review of what we've accomplished thus far, so we can be clear about what's left to do:

1. Installed CCM server.
2. Created or updated DHCP and DNS to support CCM and the phones.
3. Performed physical installation of the phones.
4. Verified phones start up correctly.

Next, we're going to perform the advanced tasks needed to make the phone fully operational. Please note that this will not be a step-by-step configuration since your needs may vary. Instead, the functional areas will be presented and discussed.

First on the list is to create what can be called *regions* in CCM. A region is an area of phones overseen by CCM, which tells them how they should communicate with phones outside their region. With phones located on the local network, using g.711 compression (or lack thereof) allows the phone to have the highest quality of voice with the least demanding processing requirements. Since the 3524 switch is Fast Ethernet capacity, using g.711 on the local network makes the best sense. To designate a region, go to the **System** menu on the **CCM Administration screen** and choose **Region** (see Figure 10.12).

Figure 10.12 Some Advanced Configuration Menu Options

When you create the regions, define them by choosing names that reflect what they are, like *local users* or perhaps *mobile users* for those using the IP SoftPhone. Our site uses two regions, named like those in the previous sentence.

Next, create device pools which logically group the physical devices. Just like regions, the names you create should define their device types. Nothing is needed to configure the device pool. It is just a logical group, like Windows NT's Global Groups or Netware's Groups, to which you can later add devices. You can then add device pools to a region so they are treated with a specific type of compression.

One other setup task is called *locations*, which is merely a definition of where a device will reside. Since CCM can handle huge numbers of users, both local and mobile, CCM administrators often use this tool to group sites by their geographical location. Let's say that CCM is located in Atlanta and serves 25 users, but Charlotte, Raleigh, Tampa, and Miami all have 5 users per site. You can create locations for all these sites, which give you the ability to control how these users access the system.

Next, go into the CCM system configuration and define the range of phone numbers the site will use. The default is the range 6000 through 6999, but you can add more ranges as needed. You should always, however, be cognizant of other site needs, so you don't run out of numbers, or create duplication of numbers between sites. Many CCM administrators reserve the 6000 through 6099 numbers for desktop IP phones, 6100 through 6199 for IP SoftPhones, as well as other series.

Now, go into the Device section of CCM and add a new phone (see Figure 10.13). This will let you choose the type of phone, what region the phone will reside in, as well as what phone number will be associated with the physical phone. Lastly, we need to create the users in the CCM directory who require these new resources. When a user is added, you must select a phone that will become the user's own phone.

Sometimes user/phone allocations will need to be changed—for instance, if the user gave up his office to telecommute from home. In such cases, the user would likely employ the Cisco IP SoftPhone on a laptop computer. So, instead of assigning this person a new phone number, all you'd need to do is create a new phone called a *CTI point*, (a Computer Telephony Instrument). This CTI point is indeed the SoftPhone. You can go to the original 7960 phone and remove it from service, then assign the original phone number to the new SoftPhone when the user moves out of the office. In this manner, the user never loses their number, nor their service.

Figure 10.13 Adding Devices to CallManager

Troubleshooting Problems with CallManager

If you can't place a call, there are several ways to troubleshoot the problem. One of these is to check the basics:

- Pick up the handset and check for a dial tone (provided by CCM). If there is no dial tone, the handset is not communicating with CCM.
- Check the handset's configuration by way of the front panel of the handset itself. In the network settings, make sure the phone has picked up an IP address, and that it has the proper default gateway. The settings must also identify that it has found CCM itself.
- If you can place a call to another phone on the inside, but not the outside, check the gateway. From within the gateway, run several debug options to see if the gateway is even communicating with CCM. Try to PING the CCM, and other devices, to make sure the gateway is IP-reachable. Run **debug mgcp all** to watch the communications stream between CCM and the gateway.
- Check the Event Logs in the CCM Windows 2000 server itself to see if CCM is having problems.

Though these are basic troubleshooting approaches, you can really break it down into two essential call types: internal and external calls. One quick test is to

place a call and observe the busy signal. Does it sound like "beep-beep-beep" in very fast intervals? This is called a *fast busy*, and indicates the device cannot reach CCM itself. If all phones are getting this error, then the problem is global. Rebooting CCM might fix the problem if you're trying to reach an internal number, but if the same fast busy exists when all users are trying to reach an outside number, the CCM will not be able to reach the gateway.

Using Cisco Unity Voice Messaging

Unity is Cisco's second-generation foray into the voice messaging world of VoIP solutions. UOne was a good product, and Cisco brought forth enhancements by using Microsoft's Exchange Server v5.5 with Unity. Unity is not compatible with Exchange 2000 as of the printing of this book, but Cisco is planning Unity upgrades which will make this a reality in the near future. Figure 10.14 shows the Unity main menu.

Figure 10.14 Unity's Main Menu

Voice mail has long been a product of telephony systems whether it be a large internal system or a simple answering machine. Some messaging systems are complex servers that store the voice message in a proprietary format, while others use a simple magnetic tape. Regardless of how it is done, voice messaging has long been an integral part of corporate life. This section briefly discusses how Cisco Unity integrates into the overall AVVID solution.

A Word about Exchange Server v5.5

In order to install and manage Unity, you must know a moderate amount about how to manage and troubleshoot Exchange Server v5.5, as well as how to have Exchange patched to Service Pack 3. Exchange uses standard directory services, while Unity uses these directory services for its own user directory. The Unity administrator will need full administrative access to Exchange in order to administer the full Unity system.

In large organizations, Exchange designers often use a design that has field mail servers connecting to, and delivering mail to, one central server, called the bridgehead server. All Exchange servers receive and deliver mail through this bridgehead, and user accounts are controlled via this bridgehead. The actual user mailboxes normally reside at the field server level. This is where Unity functions, not at the bridgehead. You should ensure each Unity user has a working Exchange mailbox at the field level, which is where the IP phones reside as well. This is also true for IP SoftPhone users since all resources must be locally attached.

Installing the Unity Messaging System

Unity is actually an add-on product to Exchange Server, yet it is installed on a separate physical server. This ensures that using the existing directory services causes Unity to retrieve current user account information to create the voice mailboxes. Because this is a separate Exchange Server from the existing Exchange Server, you'll need to ensure that your licensing is up to date for all users.

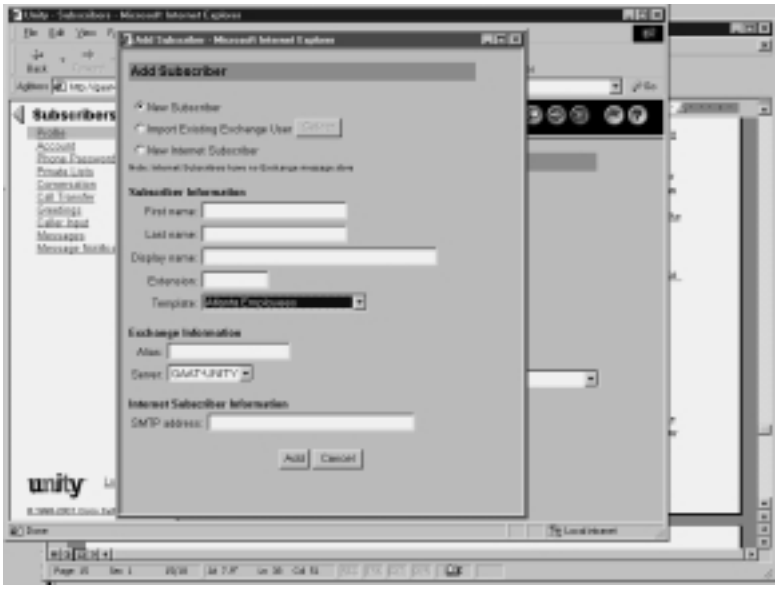
Unity comes as a CD package, so you'll need to ensure that Exchange Server is installed on whatever server you purchase for Unity. The server should be equivalent in capacity and power to a normal Exchange server. However, Unity stores the voice messages as .WAV files, which can be quite large (over 1MB) for a message several minutes in length. The actual installation process for Unity is almost completely automated; you just have to answer a few simple questions. The fun really comes when we get to the configuration section that follows.

Creating Unity User Accounts from Exchange Server's Mailboxes

Adding a new Unity user is very simple to do. From the **Main menu**, click **Subscribers**, then the + (plus sign) icon in the upper-right corner of the menu. A screen should appear, offering you choices as to where to get new user information:

- **New subscriber** This is one in which you manually type in all required information, including which Exchange server will host the user's voice mailbox (see Figure 10.15).

Figure 10.15 Creating a New Unity User



- **Import Existing Exchange User** This option searches the Unity Exchange Server for the directory services, and presents you with a list of available new users. If the user is already in Unity, they will not appear on the list.
- **New Internet Subscriber** If a person did not have an Exchange account, but still needed a mailbox, you could use their SMTP Internet address as the destination for the .WAV message files. Note that if you create this user, and they have an existing Exchange account anywhere in the Exchange domain, this new account will disrupt the user's Exchange mail delivery.

This is the simple version of installing and configuring Unity. There's really not much more to it. Sure, there are small issues, like installing and configuring Exchange Server itself, but any capable Exchange admin should be able to make short work of this.

Designing & Planning…

Do You Really Need Unity in Your Life?

In our world today, there is a constant urge to stay in touch. Cell phones, satellite pagers, even mobile television systems on campers keep us constantly in the loop. This then begs the question: Is voice mail really all that important to your site?

Before answering, first consider what resources are needed to support Unity. There's a new Microsoft Exchange v5.5 mail server, not to mention the Unity software. Lastly, you'll need an existing Windows NT domain and DNS server to support the Unity Exchange server. To run all this, you'll need an experienced Exchange administrator as well as someone who knows NY domain management and DNS services. Though this sounds like you'll need two specialists, you might be surprised to find your existing network engineer knows enough to handle all these tasks.

Keep in mind that an investment in VoIP (and AVVID solutions in general) represents an investment in people as well as technology. When you jump in, there's no climbing out without a sizable loss of capital investment. Still, the benefits of having your voice messages e-mailed to you anywhere in the world just might make your life cheaper and easier to manage, which is exactly what Unity can do for you.

Using the IP SoftPhone

Those of us that travel a good amount do not want to be left out when it come to the fun and ease of having Voice over IP. The software version of the product is Cisco's IP SoftPhone. This is a Windows-only product that duplicates the view and capabilities of the Model 7960 IP phone.

Introducing the SoftPhone

The SoftPhone requires a Cisco license for each user, just like when you purchase a Model 7960 desktop phone. The product is a 24MB self-extracting archive file that installs the phone software, the telephony drivers, and the software version of a DSP for the voice compression and CODEC functions (see Figure 10.16). The SoftPhone contains equivalent functionality, such as conferencing, hold, transfer, and call parking.

Figure 10.16 The IP Soft Phone, a Replica of the Model 7960 Desktop Model

You can install the product on any Windows 9*x* client, Windows NT v4 Workstation, or on Windows 2000 Server or Professional. The product has essentially the same internal configuration as the desktop phone in that it needs to know the IP address of CallManager, the user account that connects to CallManager, and what lines to use when the connection is made. Before installing the product, you must be sure you have all this information on hand.

Installing SoftPhone

The SoftPhone installation takes only a few minutes. Even then, you should already have a connection to the network where CallManager resides. If you don't have this connection, you won't be able to validate the immediate installation, which will cause you to have to reboot the computer to implement the changes. If you so desire, you can have the SoftPhone start up automatically whenever your computer starts up.

Configuring SoftPhone

When you installed Softphone, you were asked for numerous pieces of information that form the core of the functionality, such as where the CallManager resides so the firmware can be downloaded to operate the phone. Once these configurations are completed and a successful connection has been made to CallManager, you won't be prompted for these settings again. You have to make your changes manually the next time you want to do this, and selecting the line to use is among those changes. To do this, click the **Settings** button on the

phone, then the **Advanced** button. The phone will then attempt to connect to CallManager and determine what lines are registered to your user account (as shown in Figure 10.17).

Figure 10.17 Choosing the Line to Use

There's a configuration setting in the Telephony PBX drivers that asks if the phone should poll for new lines when it is started up. This is a double-edged sword. If it searches and finds nothing, the phone starts up using the last successful settings. If it searches and finds new lines, you'll be prompted to choose which of the new lines to add to your phone. The issue here comes from having a mal-configured CallManager, where a mobile user is attached to the network via a slow dialup connection. If the connection is terribly slow, less than 14.4 Kbps, SoftPhone may slow while trying to establish (or re-establish) a connection to CallManager. For the most part, mobile users will only use one line, so it's recommend that this setting be left unchecked and empty.

Troubleshooting SoftPhone Issues

As good as the phone is, it's only as reliable as the computer it's installed on. It is a software application with special needs, and problems do happen just like with any other application. Softphone, of course, depends upon the presence of CallManager to do its job, and as a result, most problems will arise in regards to how Softphone makes its connection to CallManager. Figure 10.18 shows the telephony settings required for Softphone to make a successful connection to CallManager.

Figure 10.18 Ensuring Proper Telephony Selections Are Made

These settings include the IP address of CallManager, the username and password for the user's CallManager account, and requesting that Softphone poll for new lines. Also, any changes made to the Cisco PBX Telephony section shown in Figure 10.18 will require you to reboot the computer for the changes to take effect, even though there is no warning or notice to that effect.

Using AVVID Applications in Video Single Site Solutions

This section details Cisco's newest solutions for dissemination of information to a larger audience: conferencing. Cisco uses two types, IP Television (IP/TV) and IP Video Conferencing (IP/VC) solutions, to exchange information between users at the same site, multiple sites, and large campuses. The information presented here, while particular to a small or single site, must diverge to serving users outside the immediate site in order to increase the benefit and return on investment. Therefore, we'll present a review of network design issues particular to these two solutions. Keep in mind that previous discussions about the VoIP solution will apply here as well, so these situations will be re-addressed here as might be warranted.

Designing the IP Network for Multicasting

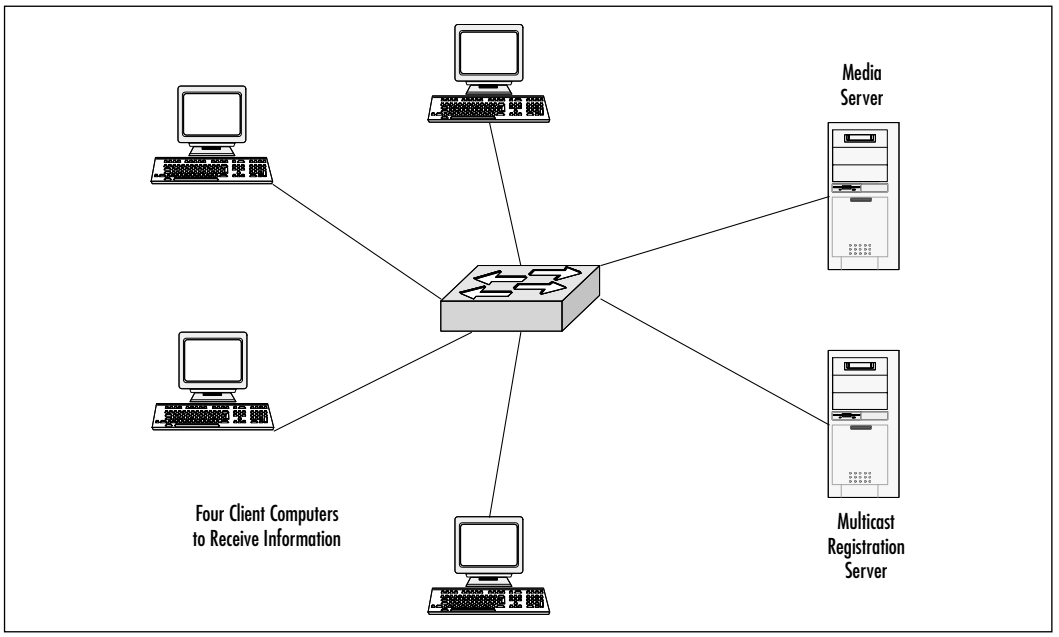
Whether you're designing a new network for AVVID, or modifying an existing one, several requirements must be met if the solution is to work correctly. Among

these is the configuration of the routers to handle multicast IP traffic. So, what is multicasting?

Multicasting is the ability of one host to transmit data to many recipients without having to address each one individually. Let's say computer A needs to send information to 25 destinations. If we looked at this as though it were e-mail, we'd have to place 25 e-mail addresses in the message's TO: line. Instead, let's create a distribution list containing the 25 recipients, each of whom will receive a proposed e-mail. The distribution list then becomes the registration point for each member. Granted, users must be manually entered into the distribution list to become a member, but the list makes life easier when sending e-mail to large numbers of recipients.

Multicasting on a network is very similar to this concept, except that computers are the recipients of the information, or more correctly, the application running on each computer is the recipient of the multicast stream. Your first impression might be to presume this is the same as standard television, but it is not. TV is a broadcast medium, meaning that everyone receives the data that the head end sends out. True, your end might be blocked from seeing channels you don't subscribe to, but your end is still receiving all the channels. At this point, however, the parallel diverges quite drastically. Figure 10.19 shows the network we'll build on for the remainder of this chapter.

Figure 10.19 Basic Network for Multicast Applications



This is a simple network with four client computers, one media server containing the streaming data to be disseminated, and what is called a multicast registration server (MRS) where the client machines must register to receive the data. This MRS can be a physical server, an appliance-based device, or a combination of the two. It is where the media administrator creates the title of the distribution application, as well as a list of the users who either want to, or need to, view the title. When this list is completed, each end recipient will receive notification of this data. The end user will then send an RSVP stipulating whether they wish to receive it or not.

The clients are running the application that will view the data, and each must register with the MRS in order to receive the data stream when it is presented, which now adds their computer to the distribution list. At the time of the showing, the data is sent out only to those who have officially registered for the playing of the data. Because we previously configured the network for VoIP quality, this means all of the AVVID family is now covered for QOS needs.

Since Figure 10.19 is a simple four-client network running at speeds ranging from 10 Mbps to 100 Mbps, all four should receive the data at full stream and get a good presentation. Problems only pop up when there are many more segments to the network, more routers, and varied speed link connections.

Local Area Networks

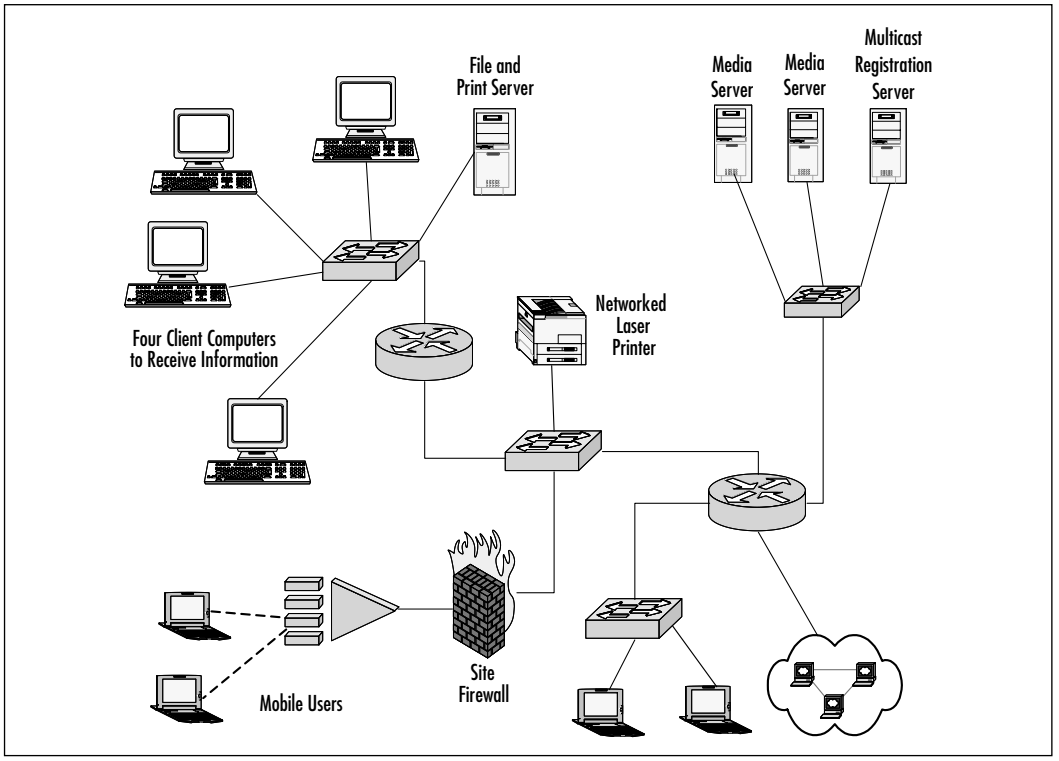
Because Chapter 10 speaks to small and single sites, multicast applications generally present no problems or issues since all servers and users are on the same network. But, as Figure 10.20 shows, single sites often grow to have multiple routers, as well as remote or mobile users. Still, on a local network, there are issues governing Quality of Service (QoS) for streaming media that must be addressed. Please refer to Figure 10.20 for the next discussions.

Here, you still have a local network solution, but you also have a truer representation of a 50-user single site solution. You now have mobile users on a dialup in the lower left corner of Figure 10.20, you have the media servers on their own network for protection and security, and, lastly, this site has its own internet connection. At the lower center of the figure, you have mobile users on laptops that come into the office on occasion, but don't have a designated cubicle. These users are possibly contractors or guests that need limited access to the network, not to the main data servers.

The main point to understand is that even with routers in-between the servers and users, everything is still on the same local site. All major link connections are

running at 100 Mbps Fast Ethernet, meaning there are no apparent bottlenecks which might cause a loss of data.

Figure 10.20 The Same Small Site Expanded



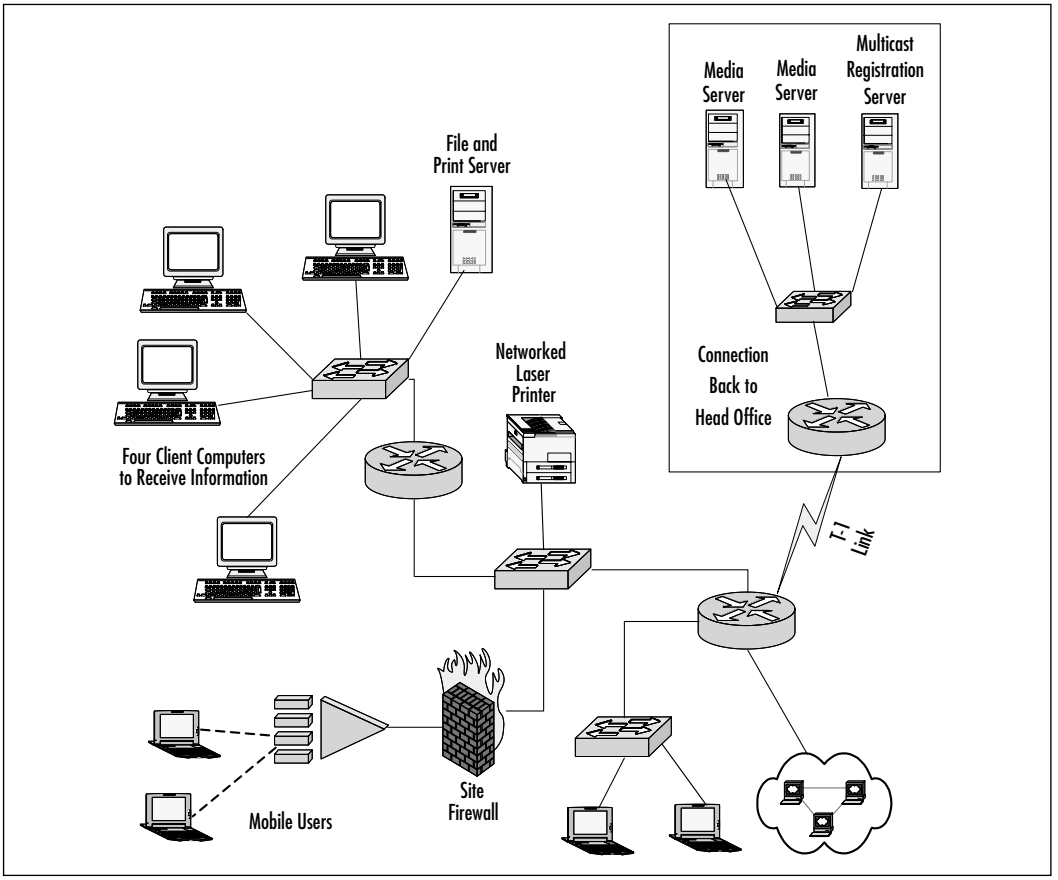
To presume this, however, would be a mistake. In this more accurate scenario of a true network, the speed of mobile users dialing into the network will be no faster than 52 Kbps, so they will need the most consideration when it comes to bandwidth. Also, you now have printing, file access, and security needs on the routers that will place router CPU processing demands upon each portion of this network.

At every part of this network, QoS must be implemented if the streaming data is not to interfere with normal network operations. Creating VLANs on this small site is not a valid solution since there are only 50 users, but by using the proper size Cisco router, you can configure each router and switch to support the desired QoS controls that will prevent loss of data when these presentations occur.

Wide Area Network Considerations

Taking Figure 10.20 to the next level, we'll presume that this site is a branch office connecting back to the corporate backbone. When this is done, the media servers usually reside at the headquarters and the data stream is pushed out to the sites. Though this is typical form, it's not the only way it can be done. Since corporate communications tend to be large, the optimum location for the media server is back at the head office, preventing a clogging of the WAN circuit out to this site. Figure 10.21 shows the media servers relocated out to the corporate network, with all remaining site characteristics staying the same.

Figure 10.21 Connecting the Site Back to the Head Office



While the local network remains the same, now the media streams must come from the corporate network across the WAN circuit. Since most streams require at least 512 Kbps of bandwidth, this now raises all manner of concerns for

the small site. Chief unto these issues is that the site may not be able to get that fast a circuit and, therefore, be relegated to a much slower xDSL connection.

It is exactly this problem that caused Cisco to devise a better schema of controls with DSL circuits for handling multicast applications. The Model 1750 and 2600 Series routers now accept ADSL WAN adapters, and the IOS firmware has much better QoS controls for assuring packet delivery without error. We know from experience that the fastest and most direct connection between the two locations is the best solution; we also know this is never possible with all sites.

So, what does this mean to small sites that still need access to multicast applications? Since xDSL is a single site solution for Internet access, these sites will most likely create a VPN connection back to the head office's firewall, which then hands them off to the internal network. This is not the best solution since it depends on the stability of the public Internet, but in the absence of a frame relay or leased line connection, xDSL coupled with a VPN currently remains the most viable WAN solution for remote and branch offices.

Remote Access Solutions for Video Presentations

Remote and dialup mobile users remain the last bastion of technical solutions for AVVID applications. Even the IP SoftPhone previously discussed required the usage of compression and advanced hardware tools to provide quality voice presentation. Add to this the same high level of quality for video, and you have a daunting task ahead.

The least problematic of these is to use audio only presentation, which was achievable in VoIP solutions. The presenters can prepare a slide show of the event so the audio-only presentations follow along with the voice presentation. Until analog dialup solutions find higher speeds, this area of frustration will remain one of the last sticky points with travelers.

There is one solution currently being tested at various sites in the U.S. and Europe: wireless Ethernet. This 11 Mbps solution places the mobile user on a higher plane of communications, albeit at a higher cost, by using the medium for Internet connectivity instead of xDSL. If the performance proves acceptable in tests, wireless communications could easily surpass xDSL as the next Internet solution of choice.

Not to be outdone, satellite systems place Internet access of up to 400 Kbps at remote and home-based sites that are sufficient to carry both the audio and video presentations, plus the VPN access required to get back to the head office. Even when the increased cost of satellite systems is inserted into the ROI equation, small sites with slow copper access may find this an attractive solution for

remote and mobile users. While these wireless solutions are usually geared to sites, network adapters for laptops and desktop computers alike are being tested that provide equal speed access to the Internet.

Cisco's IP Television Solution

In keeping with the AVVID model, Cisco's IP/TV solution is one member of the Cisco Content Delivery Network solution set. As you're well aware, television as we know it consists of hundreds of channels of programming to choose from, but you must subscribe to channels to receive them. Television is a broadcast medium, and uses various mediums to receive the information, such as coaxial cables and satellite dishes.

IP/TV differs from this only in that the medium is Ethernet networking and uses the IP architecture for the distribution medium. Cisco's IP/TV solution uses numerous devices to create the content delivery portion of the network. Looking back at Figure 10.20, the media servers are now expanded out to devices with specific functions. Before these can be discussed, let's do a quick review of public television so we have a basis for the Cisco's IP/TV solution itself.

Public television uses broadcasting to send out the data to subscribers. It is a form of multicasting in that only subscribers are permitted to see the information, but it uses a broadcast medium to carry the video and audio. At the source, a device called a *headend* communicates with the satellites to first receive the data. This information is carried in real time from the source (the television studio) down to the local sites, usually a city. This data is a one-way exchange of information towards the subscriber, and is a live feed. Think of the evening news you watch when you get home from work, and this is what you'll see.

The next type of television concerns the broadcasting of previously recorded shows, which is still a one-way exchange of data from the studio to the subscriber. There's no live person showing up on the screen, just a tape or DVD being played, or some other form of stored media being played back. After you've had supper in the evening and caught the news mentioned in the previous paragraph, you might kick back in your easy chair and watch a movie. This is the second form of television we're accustomed to seeing. The demands on the network haven't changed since the presentation is still one way towards the subscriber from the source.

Lastly, we sometimes watch public television where a talk show allows guests from other cities to interact with the main program in order to give feedback or ask questions. Placing the highest demands upon the broadcast network, this

scenario uses two-way video and audio, which requires massive amounts of bandwidth on the broadcast medium.

Uses for IP/TV

The Cisco's IP/TV solution addresses all three of these functions with the Content Delivery Network (CDN). In the order of least to most demanding on the network is the following Cisco IP/TV solutions:

- **IP/TV itself** This is a one-way multicast solution in which live or prerecorded video and audio is transmitted to the registered users of the system. An excellent use for this is executive presentations to branch offices (such as quarterly updates) to people in the field. While the presentation is being carried out, viewers can send questions to the presenter by way of a text window at the receiving site. These text-based questions can then be answered by the presenter as needed. Cisco uses this method to host monthly technical lunch meetings to inform the field about new Cisco products, and customers are often invited to participate in these meetings. If your organization subscribes to it, you could receive direct feeds from networks like NBC and CNN and display it on the client computer in real-time audio and video. However, this form of distribution consumes close to 6 Mbps of bandwidth on a continual basis per computer. If 50 users were receiving the broadcast, 300 Mbps of bandwidth is not needed—just the 6 Mbps Ethernet stream. WAN circuits should start off at a minimum of 128 Kbps available bandwidth since routers and processing devices address each CDN packet to all devices at the multicast address.
- **Ad-hoc presentations** If the previous presentation was missed, or had to be reviewed by new employees after the initial presentation, this method of delivery allows for content on demand at the pace of the user. However, this is a one-on-one media stream from the media server to the user, and is not a multicast presentation. If dozens of users activate this type of presentation at the same time, then the network would experience a sizable jump in utilization without notice.
- **Two-way multicast videoconferencing** This form of content delivery is a two-way exchange of information, much like option 1 shown earlier. However, the difference is that the presenter and the receiver can see and hear each other from all locations. This places the

heaviest demands upon the network of any delivery mechanism. Even though Cisco's IP/TV solution can do this, it is not optimized for it. The next major section on IP/VC will discuss two-way conferencing.

This solution is great for distance learning as well. With a little work, you could present lunch-time workshops and briefings to just about any site in the organization, including mobile users. With a little HTML programming, it would be very easy to create an intranet Web site listing all these CDN presentations. This would give all users of the system the opportunity to choose to view the presentation at their leisure. In the next section, you'll see how Cisco devices perform this little bit of magic.

Devices Used in IP/TV Solutions

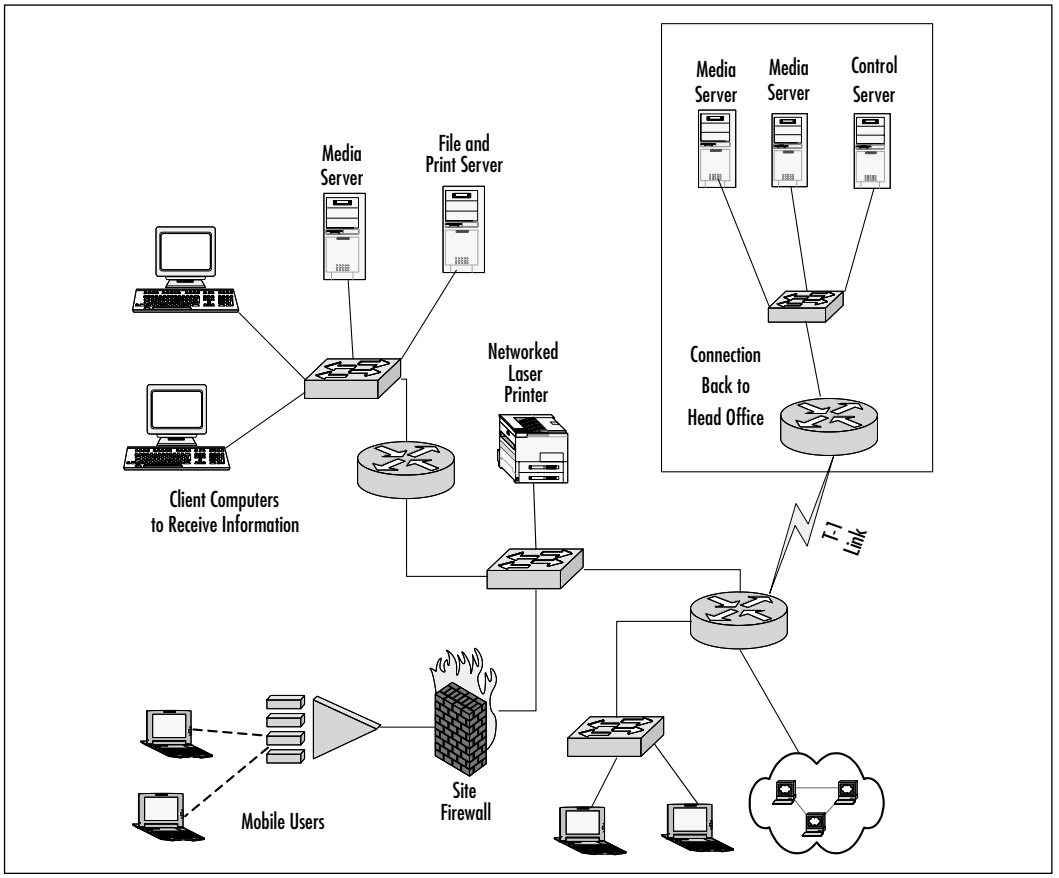
There are several devices in Cisco's IP/TV solution that make this possible:

- **The IP/TV Control Server** This device constitutes the registration and policy controls for the system. It handles coordination of the broadcast, encoding of the source (such as that from a video camera), and control information for the IP/TV client software installed on the desktop computer. This server is capable of assuring that only the desired clients receive the broadcast, and can encrypt the broadcast if need be.
- **The IP/TV Broadcast Server** This server handles the actual processing of the streaming media, and distributes it to the destination based upon the control information hosted in the Control Server. Each Broadcast Server processes only one stream format per session. If you need to present in MPEG-1 and MPEG-2 formats simultaneously, you'd need one Broadcast Server for each media type. The Broadcast Server accepts source input from many locations, including live feeds, a VCR tape, DVD, and live recording in a local studio.
- **The Archive Server** When broadcasts must be played back as an ad-hoc presentation, the source information comes from this server. A one-hour full video presentation in MPEG-2 format can easily consume 45MB of storage space, so this server should have suitable storage capacity and flexibility to grow. You could also use a Storage Area Network to house these source files.
- **The IP/TV Client Viewer** This is the software application that runs on each client computer that will receive the broadcast, and effectively

completes the multicast registration to the Control Server. This registration uses the media access control (MAC) information on the network adapter as the method of assuring proper delivery.

Figure 10.22 shows how you might have a combined deployment of Cisco IP/TV solutions both at headquarters and at branch offices.

Figure 10.22 The Final IP/TV Solution



Single Site Solutions for IP/TV

Earlier, Figure 10.20 illustrated a typical small site with many users and network resources. The optimum solution for these small sites is to have a broadcast server located at the site with a connection back to the corporate control server. When presentations are required that might be particular to the users at that site, the

head office could distribute the presentation to the affected sites late at night so site connectivity is not affected. Users can then view the data without affecting performance of the WAN link or the Internet connection.

This also offers flexible connectivity solutions to mobile users closest to that site, since they're more likely to be able to come into the site to view the presentation without dealing with the problems of a dialup analog modem connection. This isn't to say that users at this site must have a broadcast server on the premises, but it might be a better alternative than paying for faster and more expensive WAN circuits.

Cisco's IP Videoconferencing Solution

This last section of the chapter will extend the IP/TV solution just presented in the previous major section. The same principles exist for the QoS needs of the network LAN and WAN sections, but this should not change once established in the network design. What *will* change is that this two-way exchange of data now potentially doubles the bandwidth requirements on the system. This section will discuss these issues as they relate to a small site's needs.

Conferencing is not new. It has been around for some years under the H.320 standard, and works across ISDN quite well at 128 Kbps bandwidth. Early adopters of videoconferencing experienced latency and frame loss that was quite unacceptable, but this improved with new encoding standards. The latest generation of videoconferencing now uses the H.323 standard of call processing, which includes VoIP solutions. Typical solutions for Cisco's IP/VC solution include:

- **Distance learning** This permits the two-way exchange of information between instructors and students through the use of small desktop cameras and microphones.
- **Medical consultations** Doctors across the country and around the world can now conference in experts and residents alike to find quicker solutions to problems without incurring the expense of travel. More importantly, medical diagnosis can be completed in minutes by sharing X-ray information between sites as the emergencies occur, in some cases very close to real time.
- **Financial institutions** Imagine being able to work with your bank's customer service department to apply for a loan or reconcile an account

issue, all without leaving your home. This is already in service now with several large banks.

- **Training** These AVVID solutions are not the easiest to learn and manage some times, so why not conduct personalized training of networking systems as the need arises? IP/VC permits scheduled, ad-hoc, and spur-of-the-moment conferencing to any IP/VC-enabled locations, provided the bandwidth is available.

This section will talk about these topics and how you can leverage them in your organization, especially with small sites in mind.

Equipment Uses in IP/VC Solutions

Cisco's IP/VC solution uses a number of devices to achieve their goals. Cisco solutions are backward compatible with the aged H.320 standard to help customers migrate to the latest standard without a loss of investment. Cisco's solution uses the following devices, although not all of them are used at each deployment:

- **Multimedia Conference Manager** This is an IOS-based gatekeeper solution that provides call admission control, proxy management, call detail records, and security controls for each VC connection.
- **IP/VC 3510 Multipoint Control Unit** Serving as many as 15 connection points, this device provides gatekeeper controls for small sites with speeds up to 768 Kbps.
- **IP/VC 3520/3525 Video Gateways** These two units provide BRI and low speed (384 Kbps) connection video processing of H.320 to H.323 schemas. The 3520 supports BRI capacity while the 3525 supports a single ISDN-PRI connection to legacy systems. Both units provide internal gatekeeper functionality.
- **IP/VC 3540 Multipoint Control Unit** This unit is much the same as the 3510 unit, but it supports up to 100 connection points as well as the controls and increased processing capacity needed for such a high number of connections. It also supports connection speeds up to full T-1 capacity.
- **IP/VC 3540 Application Server** This Windows NT server is used to provide applications level support for whiteboarding and applications sharing.

- **IP/VC 3540 Gateway Module** This piece of hardware provides H.320 to H.323 conversion up to 384 Kbps connection speeds with the most popular video types and formats.

While not an exhaustive list, these are the major devices used in creating the actual IP/VC network. You could easily use both the IP/TV and IP/VC solutions on the same network provided that adequate QoS controls and WAN connectivity are in place.

Good Examples of Using IP/VC for Small Sites

Presuming that the bandwidth is there to properly support Cisco's IP/VC solution, there's a nice way to use IP/VC at any small site, and that's where we mentor one another. Impromptu training sessions are a boon to productivity when you can have a friend help you understand how something works, and learning AVVID is no exception. The Comptroller at the head office could assist the site financial advisor with new accounting procedures. Sometimes, simply talking over the phone is not enough. Being able to look at and touch the problem can more quickly solve it.

We all like to think we know our jobs well, and are good at what we do. This author has experienced days of wisdom followed by nights of sheer terror, especially when he's had to train his replacement. I know others who have been in the same position. Once, when I was promoted to a senior engineering position, I had to interview and train my replacement before I could leave. The problem was I was at our headquarters' main data center, but the interviewee was at a branch office. Thankfully, our company's videoconferencing capabilities solved the problem.

Have you ever had to ask for driving directions to a location before getting on the plane to fly there? Why not just whiteboard it on an impromptu IP/VC session and get it right the first time. Too few people think outside the box when it comes to uses for IP/VC, even though its uses are incredibly diverse. Before chatting to someone over the phone about work issues, imagine what it might be like, using IP/VC.

Why IP/VC May Be Bad for Single Sites

We're not bashful in giving a thumbs down to any solution (regardless of vendor) that doesn't fit the problem. As in VoIP solutions, IP/VC requires a commitment to the system once the decision is made to install it. Capital investment and periodic upgrades may be necessary to keep the system at top performance. It then

follows that the commitment must be made to the TI staff who must maintain it at every site in the system.

Likewise, past methods of doing business must be evaluated regarding how you did business, and why you want to do it now with IP/VC. If someone were to ask you if IP/VC was useful to you, and you couldn't immediately answer with a resounding "Yes!" then you're not yet ready for it. Just because it's the rage of the technology world is no reason to jump into the pool with everyone else.

An equally disconcerting issue is the recent fallout of the Dot Coms. The inability to find an absolutely stable ISP would be the death knell for any AVVID solution. Without the growth ability of the ISP link, this would also be highly damaging to the AVVID solution. Even if you discount the support issues of the internal equipment, the ISP will likely be unreceptive to your requests for help in troubleshooting any connection problem, especially after they hear you're deploying your own AVVID solutions.

Summary

Chapter 10 explored the usefulness of VoIP and a few AVVID solutions in a small site environment. You've seen how sites can be designed from scratch, and modified from an existing infrastructure. These two options contain many issues of merit, including addressing the usage of mixed vendor network electronics within the same Ethernet network.

CallManager is the IP version of a PBX system. You got a brief preview of how to install and configure CallManager, how to use it for flexibility purposes when creating and managing IP phones, and how your mobile users can still benefit from VoIP by using the Cisco IP SoftPhone. Since VoIP mandates a reasonable return on investment to justify the capital expenses, we discussed how you can gain significant ease of administration (both locally and remotely) and flexibility for handling short notice phone changes.

Lastly, we showed how Cisco's AVVID solutions are a boon when distributing information to a large audience by using the IP/TV and IP/VC solutions, which can significantly reduce travel expenses and shorten the time to deliver presentations. You can host very short notice meetings, whiteboarding discussions, and one-on-one conferences with very little cost beyond the initial installation and testing.

In Chapter 11, we'll expand on these topics to cover multiple sites and a larger campus enterprise.

Solutions Fast Track

Using AVVID Applications in IP Telephony Single Site Solutions

- ☑ Single site VoIP systems can be a cost-effective replacement for traditional PBX systems, especially in locations where available PBX solutions are limited. This is most helpful in places where you have more network engineers capable of managing Cisco devices than traditional telephony solutions.
- ☑ VoIP permits easy remote management of the entire system via CallManager's Web interface. Even the server's services can be stopped and restarted by way of the Web interface.

- ☑ By using the inline power enterprise model of switches, the customer can future-proof growth needs for both voice and data applications, foregoing the need for replacement devices and the consequent disruption of existing services.

Using AVVID Applications in Single Site Solutions

- ☑ With the development of the Unity product, Cisco provides great messaging capability that finally breaks all ties to traditional telephony systems. Now, full deployment of AVVID solutions can be achieved to other sites by using only external WAN communications, as well as all internal communications riding on the Cisco-powered enterprise.
- ☑ Because Unity integrates with Exchange Server, and uses the native Exchange directory services, it is easy to deploy and manage, and has the flexibility to handle various messaging needs. Unity works with all standards-based SMTP, POP3, and IMAP4 clients, maintaining ease of use and portability between clients.
- ☑ CallManager provides excellent flexibility for moves, adds, and changes. Its Web interface makes the system accessible from any location, even from dial-up modems with slow speeds. CallManager is highly extensible, allowing it to serve thousands of users in a centralized or distributed environment.

Using AVVID Applications in Video Single Site Solutions

- ☑ Cisco video solutions offer dramatic savings in the area of training by dramatically reducing or even eliminating travel costs. Presentations can be shipped to the site when so desired, and easily deployed.
- ☑ The flexibility to present video on demand speeds information to users whenever needed.Video on demand (VOD) means users can come back from vacation and review that missed presentation from the head office without needing to schedule a new briefing.

☑ Video solutions allow for remote mentoring at any time, by anyone. New personnel no longer have to fly to the head office for indoctrination, nor do they have to wait for the next session. Trainers can also create their own labs and exercises where the experts reside, without any travel costs. The new videos can then be shared at any location.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: What is Voice over IP?

A: VoIP is the ability to digitize analog voice signals, insert the digital data into an IP packet, and transmit it across an IP network just like any other kind of data.

Q: I've heard something called *toll bypass* helps save money. What is it?

A: Sending voice packets across an IP network does not use the public switched telephone network, and thereby avoids the cost of a public phone toll call so long as the call can be completed across the company's TCP/IP network.

Q: VoIP phones use a device called a Digital Signal Processor (DSP). What is this, and how does it work in a VoIP setting?

A: A DSP is a set of electronic circuits that collectively provide digitization of analog voice signals. DSPs also provide the mixing of voice signals such as for conference calls. Each gateway uses certain DSP modules to provide compression of the digital voice signals in order to handle different speeds of lines.

Q: I've been told I need a CallManager to run the phones. What is it, and how many phones can it control?

A: CallManager is the IP version of a PBX. This is a specialized Compaq server running Windows 2000 and the Cisco CallManager application. CallManager provides call admission and control for VoIP calls, conferences, and streaming

audio. One CallManager server can provide control for as much as 2500 VoIP devices.

Q: I've heard about people using their computer to place phone calls. How is this done?

A: They use a software application called Cisco IP SoftPhone, which installs on the client computer. When the program is executed, it looks on screen just like a Cisco Model 7960 IP phone (a desktop model) with most of the same controls and features, and uses the IP network to transmit and receive calls the same way IP phones do on the desktop.

Q: How does the Cisco VoIP solution provide voice messaging?

A: By using the strengths of Microsoft Exchange Server v5.5, Cisco's Unity messaging software integrates WAV files recorded from voice messages and stores them in your mailbox for retrieval anywhere as e-mail.

Q: I've heard that IP/TV is the same as a television broadcast. What is IP/TV?

A: IP Television is the transmission of digitized television across the IP network as IP packets for distribution to the intended recipients as a multicast data stream. This means that a Cisco-powered network using the proper devices can function as a TV distribution system as well as a Voice over IP solution.

Q: Is this IP/TV solution the same as videoconferencing? What is IP/VC, and how does it differ from IP/TV?

A: IP Videoconferencing is the transmission of two-way television data, much like IP/TV, but in both directions so the sender and receiver can see and hear each other. IP/VC uses the same Cisco-powered network as Voice over IP and IP/TV, so the maximum cost savings and benefit is realized.

Q: What is Cisco's Content Data Networking?

A: CDN is the overall part of the AVVID solution that uses IP/TV and IP/VC for dissemination of audio and video data to multiple locations.

Q: To use all of these AVVID solutions must require a great deal of connectivity to branch offices. How are remote and branch offices' WAN connections affected by the deployment of AVVID solutions?

A: Small and single sites must make sure their WAN connection can achieve a minimum speed of 128 Kbps in order to deploy these solutions. Sites with up to 50 users might require a minimum speed of 384 Kbps if VoIP, IP/TV, and IP/VC solutions are all deployed at the site.

Chapter 11

Designing and Implementing Multisite Solutions

Solutions in this chapter:

- **IP Telephony Multisite Centralized Call Processing Solutions**
- **IP Telephony Multisite Distributed Call Processing Solutions**
- **Multisite AVVID Solutions**

- ☑ **Summary**
- ☑ **Solutions Fast Track**
- ☑ **Frequently Asked Questions**

Introduction

In this chapter, you'll extend your knowledge of Chapter 10's single site VoIP solutions into a multisite corporate environment. We'll be performing specialized network designs geared to the AVVID functionality. These solution designs will evaluate the benefits and detriments of a centralized design versus a distributed environment for large environments, and will tackle the following design and operational issues:

- Providing cost-effective small site connectivity while providing required CallManager redundancy.
- Assuring a seamless growth path when a small site grows to consume more network resources.
- Ensuring that CallManager solutions are flexible in their coverage of the corporate users.
- Providing the network engineers and managers with adequate documentation of the design, and showing how the various AVVID solutions fit within each part of the design.

The solutions will first review what you've learned in the Chapter 10 single site solutions, then expand each of those topics out to a full corporate system. We'll show you how to build redundancy and resiliency into each design, how to build out clustered CallManager solutions. Lastly, you'll learn how to deploy other AVVID solutions in this same corporate environment. When you finish this chapter, you'll have at least a solid view of the minimum requirements of a Cisco AVVID enterprise network.

IP Telephony Multisite Centralized Call Processing Solutions

In centralized solutions, CallManager and all of the related VoIP resources are located on the main corporate backbone networks, or at some other primary location. These VoIP resources are any device or function that provide core VoIP functions for everyone else, and which usually have the highest capital cost. This is usually a data center or some other highly protected location that uses conditioned power, redundant WAN connections, and physical security such as personnel badges and magnetic entry cards. Because of this seemingly precarious

position, all infrastructure supporting this configuration must be of the highest quality, and utilize the most redundant design possible.

A centralized call processing solution is arguably the configuration most often found in enterprise VoIP solutions. This section will show you how to design such a solution, plan for WAN changes to support branch offices from a centralized solution, and how to provide backup and disaster recovery solutions that will help recover failed installations.

Wide Area Network Considerations

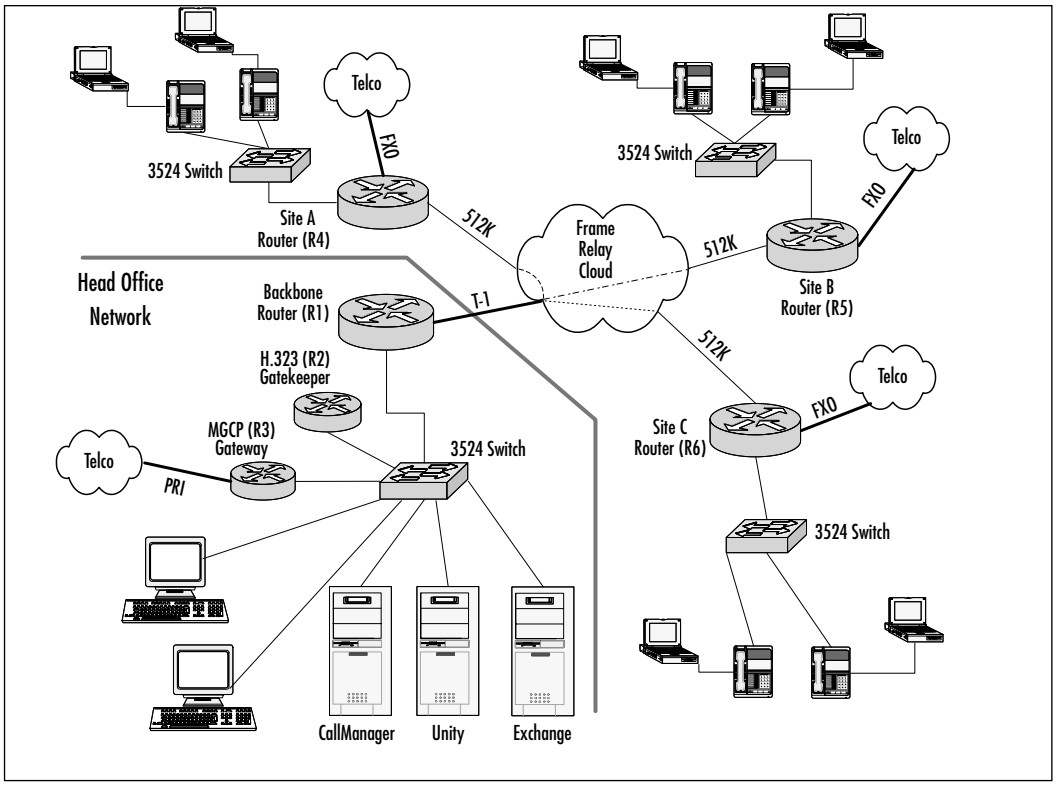
In centralized network designs, all CallManagers reside on the head office network (as do associated solutions like Unity messaging) in a central location such as the main head office backbone, not at field locations. Figure 11.1 illustrates a typical centralized design. (Figure 11.1 will be used as the main reference point in this section, and will be adjusted to reflect the amended designs explored throughout the chapter.)

This is a *balanced* design, meaning that the capacity of the WAN circuits to the branch offices equals the maximum capacity of the head office. This means that if you add up the speed of the WAN links to all branch offices, the total does not exceed the head office WAN connection to the frame cloud. Referring to Figure 11.1, the three branch offices each use a 512 Kbps connection, which totals 1536 Kbps. Since this is equal to the head office WAN connection speed, the head office WAN connection cannot be over-subscribed. This is a very important factor to consider when designing the VoIP solution.

The total VoIP seats in the branch offices cannot exceed the capacity of the circuits, nor the centralized CallManager. Off-net calls are routed to, and placed through, the Primary Rate Interface (PRI) to the telecommunications office that is local to the head office network. In this manner, head office management can negotiate the best rates for local and long distance calls, and also get the maximum utilization out of the Frame Relay circuits by using the voice and data paths together.

However, notice that the branch offices use a FXO connection to their local telecommunications office to off-net their local calls instead of routing them across the Frame Relay circuit to off-net them. The FXO ports use a standard analog telephone line instead of a specialized PRI circuit, and the cost is dramatically different. Also, standard analog lines are available in nearly every town in the country. If you can get an analog line, this is the first step towards centralized design. We'll get into this more in the section about creating the off-net solution.

Figure 11.1 A Typical Centralized VoIP Design



The Gatekeeper Function

The gatekeeper is a Cisco router that runs the H.323 MCM feature set, and provides the H.323 centralized call admissions control for the enterprise, call setup, and related management issues. Among these functions is the decision regarding whether the destination path can support the required bandwidth requirement of the device placing the call. To illustrate this concept, let's go through a call, referring to Figure 11.1 as a common reference point.

A user on Site A wants to call a user on Site C. When the Site A user picks up the phone and gets a dial tone, this person types in the digits of the destination phone. This request is sent to the CallManager on the head office backbone, which determines that the destination device is on Site C, and then contacts the gatekeeper. The gatekeeper looks at the request in regards to the amount of bandwidth requested, the type of services requested, and then makes the determination as to whether the total amount of bandwidth is available to the site.

The gatekeeper knows these things because it keeps track of the amount of calls currently placed to Site C, and the amount of bandwidth dedicated to that site. With the WAN link currently set at 512 Kbps, the average g.711 call uses 64 Kbps of bandwidth, which means that 8 simultaneous calls are possible to Site C from any other site, provided that none of the 512 Kbps is used for data streams. If other compression techniques are used, the voice streams can be compressed to as low as 5.3 Kbps with the high-complexity digital signal processor (DSP) CODECs in the voice-capable gateways.

However, Cisco design rules state that no more than 75 percent of the circuit capacity should be used for voice traffic. Furthermore, overhead in the IP packets can raise the total per-call bandwidth requirement for a G.711 call to 80 Kbps per call. Using these parameters on the same 512 Kbps connection now yields the primary reason many VoIP designs fail to meet expectations: 75 percent of 512 Kbps is 384 Kbps. Divided by 80 Kbps per G.711 call, we now have a maximum of four possible calls at the same time. This is quite a difference from the previous paragraph, and illustrates how and why these designs sometimes go wrong.

NOTE

The gatekeeper does not handle the actual voice stream between the two endpoints, but rather assures that the proper bandwidth is available between the two endpoints.

Voice-Capable Gateways

As explained in Chapter 10, a *voice-capable gateway* is a Cisco router that runs the MGCP IOS firmware that performs processing for voice calls on the local network to local or external destinations. These routers are installed with PRI, FXO, or FXS ports that form the external connectivity to a local telecommunications carrier office. The voice-capable gateways for branch offices are:

- Model 175x for small site gateways, for up to 10 users
- Model 26xx for small sites, for up to 50 users
- Mixed variations of these two devices

These two models are frequently used units; the Model 175x is the more cost-effective unit, but has less flexibility than the 26xx series and VG200 gateways.

The field gateway router used for data only might also be an older 2500 or 3600 class router that has been at the branch office for quite some time. Also, newer Model 1600 series routers may be positioned as small branch office gateways to handle the data portion of the site. It is important not to not mix up these gateways, and equally important to not try and use one gateway for both data and voice combined. While such a combination has worked at times, it usually is not a good idea to have all your eggs in one basket.

The important thing to understand is that voice-capable gateways exist to provide external telecommunications connections at that site. The nonvoice-capable gateways can still be used in a centralized environment where all calls are passed through the central site, and there are no off-net local calls. While the central site would then bear all telecommunications costs for the branch office, this isn't necessarily a bad thing. If the pre-VoIP design assessment found that 95 percent of all calls were to the head office, then the cost of the remaining 5 percent of calls could be routed through the head office backbone, resulting in that 5 percent being all long distance calls back to that branch office, but now coming from head office and not the branch office. However, you must be aware that the 5 percent of rerouted calls could substantially increase your long distance toll call costs, and thus should be a factor when deciding how to reroute calls like these.

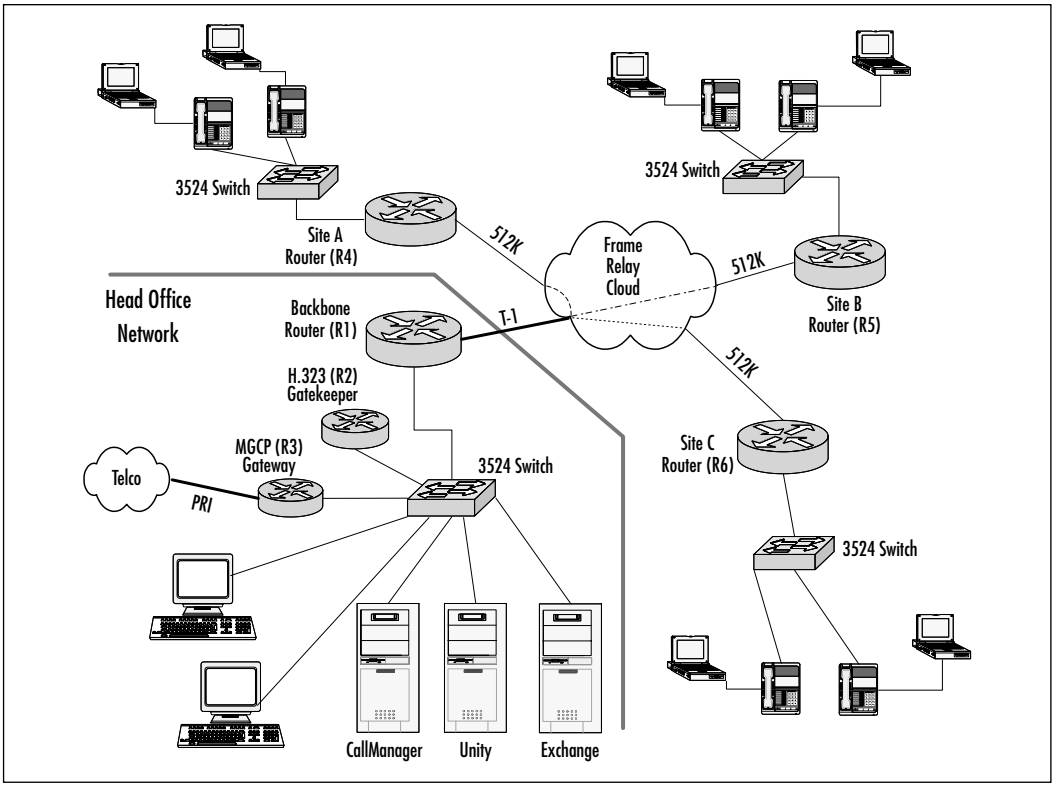
This is just one example of how VoIP solutions must be approached in any part of the design. The cost savings realized by not purchasing the voice-capable gateways might be realized in that 5 percent of long distance calls. With long distance calls now costing as little as four cents per minute from major carriers, this might just be a negligible expense. Look at Figure 11.2, and you'll see the changes in removed external telecommunications costs.

This is possible if VoIP MGCP firmware is used, but the site will not have any options to create external connectivity without replacing the router and adding the new telco cards, causing site downtime. Notice that routers R4 through R6 have no external connectivity, nor do they have a gatekeeper at each site. This is because the WAN circuit is powerful enough to centralize those functions and still carry the data load as well.

Choosing Frame Relay or Leased Lines for Site-to-Site Connectivity

The arguments for choosing Frame Relay or leased lines has caused some of the most spirited debates possible, but it must still be discussed no matter what. Frame Relay is less costly than using leased lines, yet it's usually stable enough to carry the load that leased lines do. So what's the difference that causes the cost delta?

Figure 11.2 Nonvoice-Capable Gateways Remove Extra Costs



Frame Relay uses a shared medium "cloud" provided by the telecommunications carrier. While your circuit goes from your premises to the provider, the circuit ends and hits the "cloud," so called because no one really knows (except for the provider) where the data passes through the network devices. All you know is that the data arrives at the destination safely. Figure 11.3 shows an example of a Frame Relay cloud used by many subscribers.

This cloud spans the United States and is typically joined by several telecommunications carriers. This cloud is really a series of clouds that serve specific areas of the country, and specific portions of each state as well. These connections are joined by what is called a Permanent Virtual Circuit (PVC). A PVC is nothing more than an increment of 64 Kbps channels bonded together to form the desired capacity of circuit, up to the limit of the carrier. Figure 11.4 shows an expanded view of the state of Florida to show the frame clouds at each of the major cities displayed.

Figure 11.3 A Frame Relay Cloud

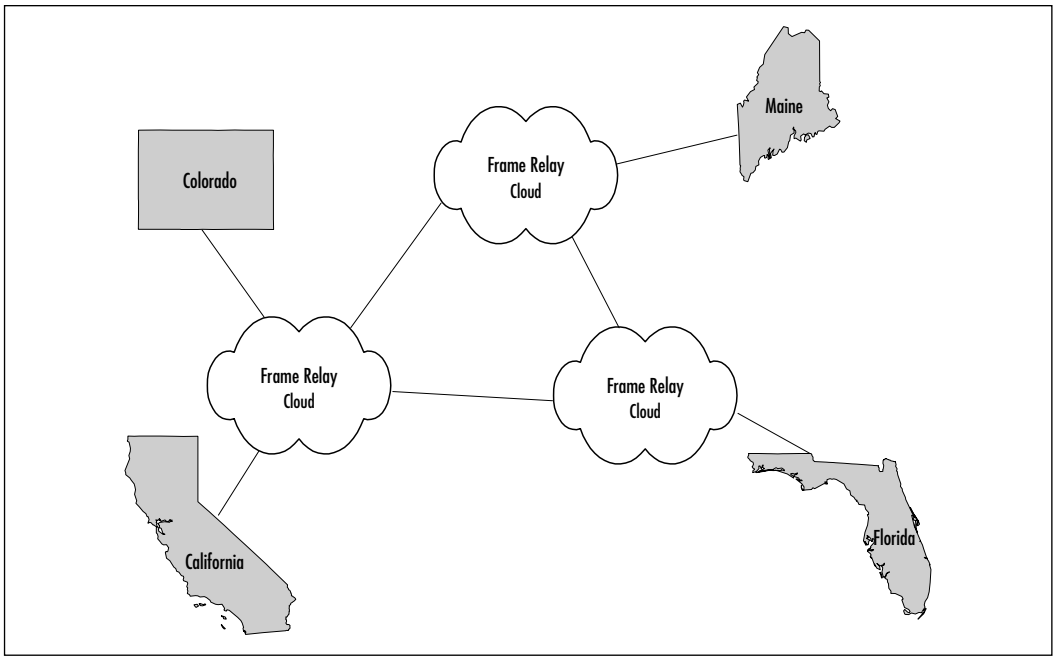
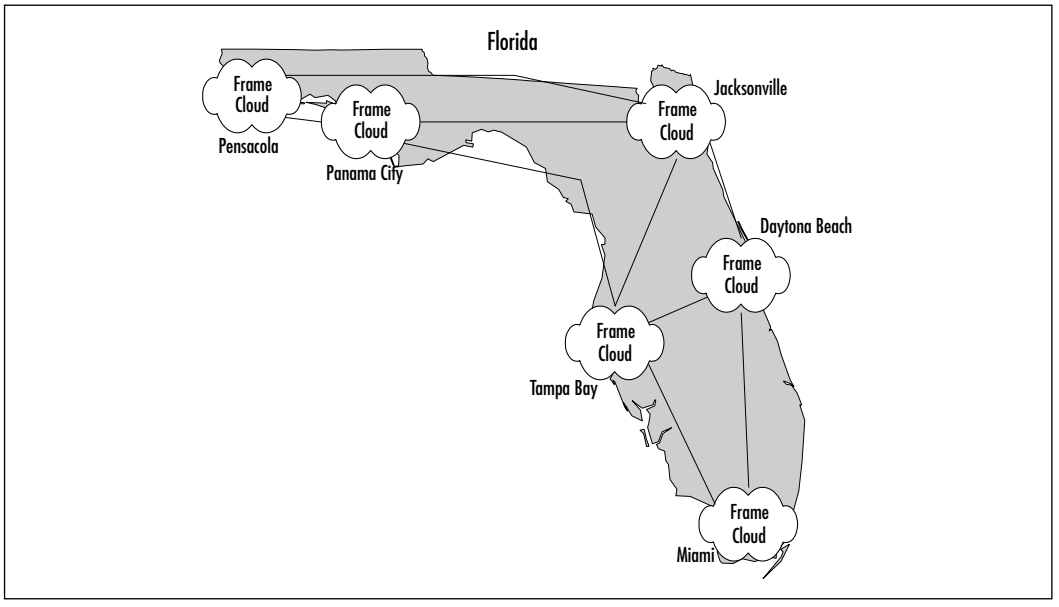


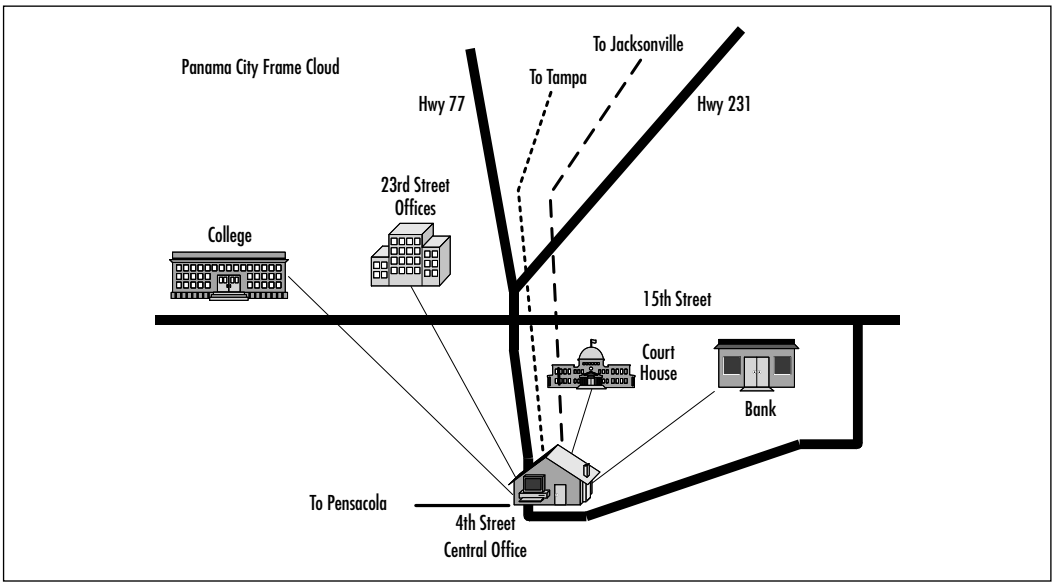
Figure 11.4 The Florida Frame Relay Cloud



This illustrates why connectivity is available in some areas, but not others. Panama City is situated on the coastline of the Florida Panhandle, whereas Pensacola sits on a major junction of highways and cities. Between Panama City and Tampa, all along the southern coastline, little in the way of major commerce exists to warrant the high cost of running the fiber optics cables required to carry Frame Relay communications. Notice how the cities are interconnected in what is called a "full mesh" that assures each city has two or more paths to take between cities. All of these circuits are the responsibility of the carrier, or carriers in some cases, to maintain and grow as demand warrants.

However, cities often expand beyond the coverage of their particular communication form (like in Figure 11.5, where Frame Relay spreads out of the central office to the businesses).

Figure 11.5 A Typical City Frame Cloud



From these series of figures, it should be clear the bulk of the risk, expenses, and maintenance sits squarely on the shoulders of the carriers. The users only need be concerned with the local connections between the central office and their location. But, when the Frame Relay cloud gets cloudier, increased traffic can impede your traffic, and cause all manner of problems. This is why frame carriers use two functions of Frame Relay to control traffic:

- **Port speed** This is the speed of the port on the router where the connection initiates from the central office, and can be as high as a T-1 of 1.536 Mbps. This is sometimes called the *burst rate* of the connection.
- **Committed Information Rate (CIR)** This is the circuit speed the provider guarantees you'll get all the time, regardless of how many subscribers are on the frame cloud.
- **Committed Burst Size (Bc)** This is the maximum volume of data the network agrees to move through the frame cloud under normal working conditions.
- **Excess Burst Size (Be)** Under normal working conditions, this is the amount of data above and beyond the Bc mentioned in the preceding bullet.
- **Discard Eligible (DE)** This is the Be data marked as lower priority than Bc data; if the frame cloud gets congested, Be data marked with its DE bit set can be discarded to help reduce frame cloud congestion.

For most customers, the CIR is one half of the port speed, so a 256 Kbps circuit would have a CIR of 128 Kbps. You pay for the CIR, and a marginal amount higher for the port speed. But, if your traffic flow exceeds the CIR, and the frame cloud is congested, then the carrier can discard your packets at its own judgment to reduce the traffic in the cloud. This means your traffic flows must slow down to account for the congestion.

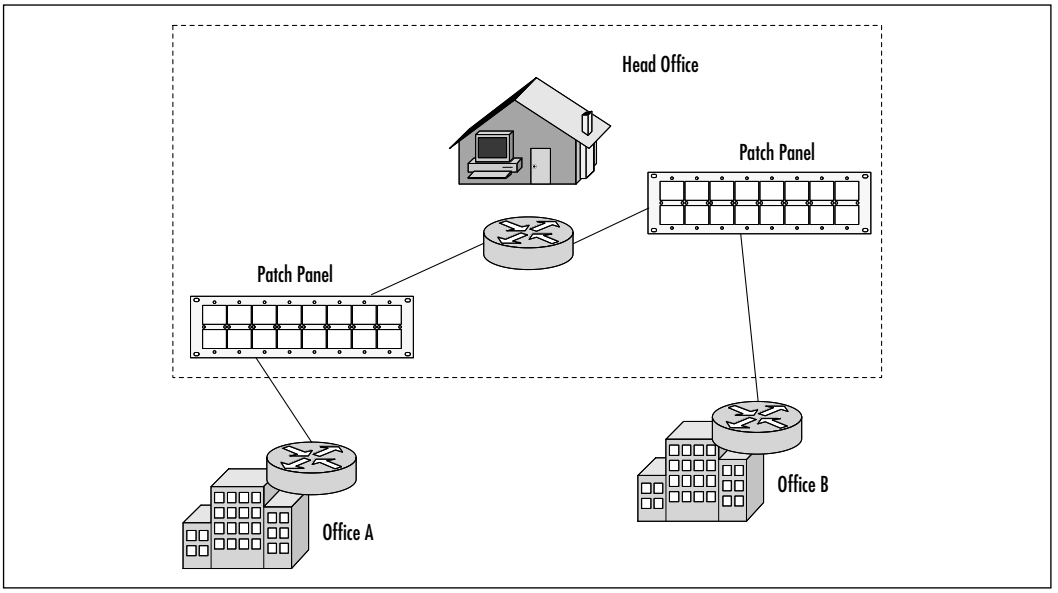
For the most part, Frame Relay works fairly efficiently. But if your connection must remain reliable and not experience discarding of packets, then your only option is to use a leased line circuit (shown in Figure 11.6).

Leased lines can easily exceed three times the cost of a Frame Relay pipe, because the connection is 100 percent dedicated from the carrier to your connection. Figure 11.6 shows two sites connected via a leased line, which is directly connected to the central office. In some leased lines, the router in the central office is a massive unit that can host hundreds of connections. This figure has been broken out slightly to show that in a leased line connection, there are patch panels between devices, but only to create the physical circuit directly between devices.

The benefit is that at whatever speed you subscribe, you get it on a constant basis regardless of the number of people subscribed to the carrier. Your connection is truly independent, but you'll most certainly pay for that privilege. In VoIP systems, if the sites are within a few miles of one another, leased lines are usually

the best way to go. If the sites are many miles apart, then Frame Relay may be the only way possible, physically and financially, to achieve the design.

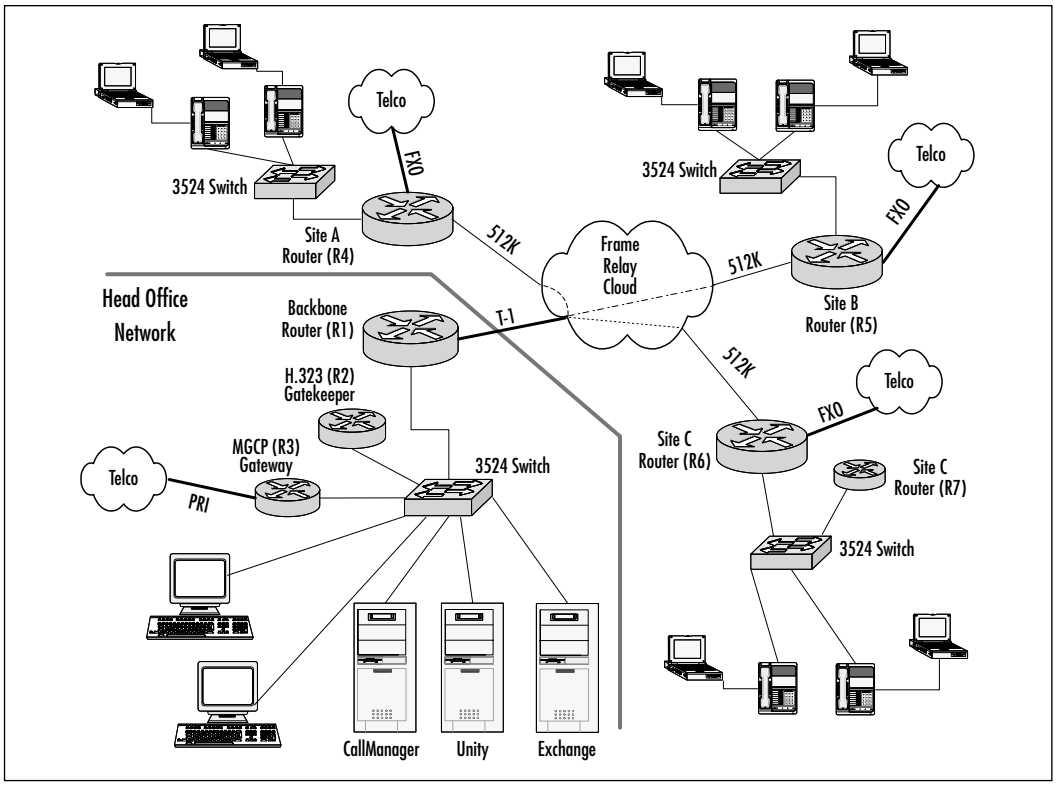
Figure 11.6 A Typical Leased Line Circuit



Using the Gateway for Data and Firewall Access Control

Chapter 10 enumerated several arguments for and against using the same branch office gateway for both voice and data processing. For cost reasons, we'll presume that only one gateway is possible no matter what. If we look back to Figure 11.1, the presumption can easily be made that if Site C had 25 users, then the composite voice and data demands upon the gateway would be truly awesome. So, a single router solution to provide that volume of power would be the Cisco 2651, since this unit is capable of supporting LAN, WAN, and voice I/O cards in one gateway. While the 175x series gateway can easily handle 25 users, it does not have the port expandability of the 2600 class gateway. But, to really isolate the data and voice functions, you could instead purchase a much cheaper Cisco 1601 (R6) to handle the Frame Relay data-only connection, and a separate Cisco 175x (R7) to handle the external off-net calls plus the voice overhead. The overall cost of these two gateways is nearly equal to the single 2651, and accomplishes the job of task separation. Only one extra gateway needs to be added to help manage the network. Figure 11.7 shows this new configuration.

Figure 11.7 The New Voice Configuration at Site C



Before you knee-jerk away from this configuration, the 1601 and 175x are more than enough to handle the tasks at that site. Since the 175x will be bound by the number of outside lines it can handle, this site is adequate for up to 15 users before bandwidth problems will occur. If more outside lines are needed than the 175x can provide, then the 2600 Series gateway becomes necessary. This is why, despite the increased cost of a single gateway solution, future growth may dictate which device is used.

Handling LAN Problems for Multiple Sites

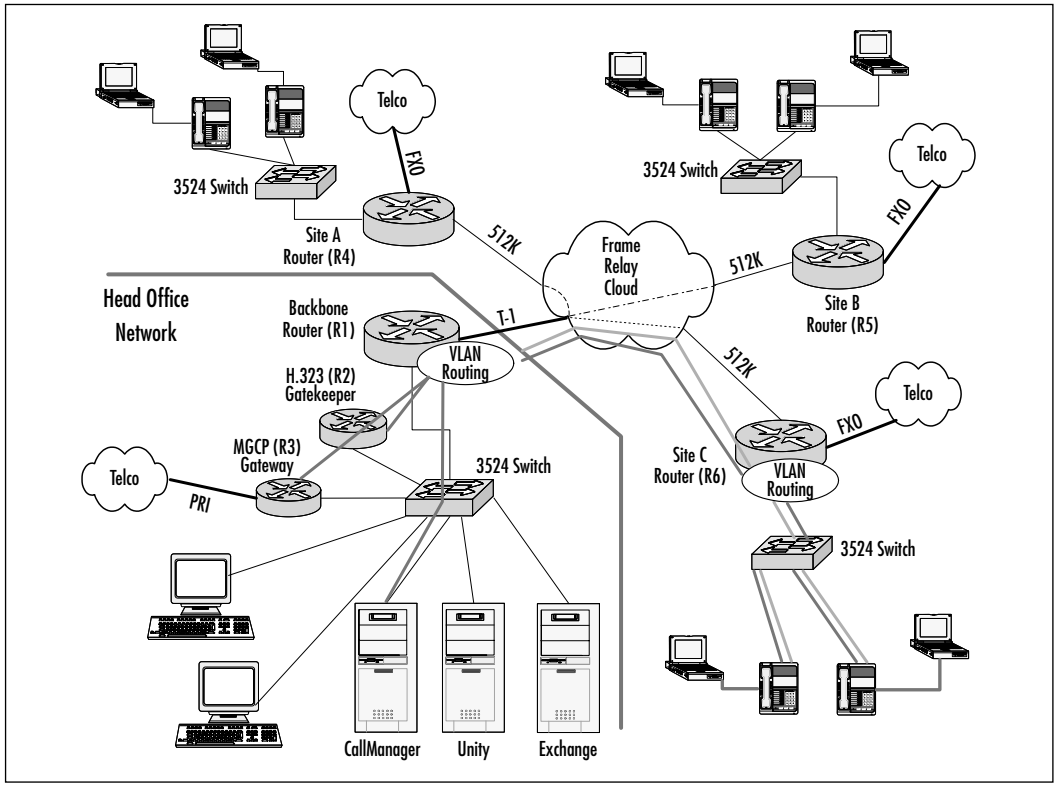
Having multiple sites converge on the head office LAN might present problems with the routing and processing of calls, especially when the centralized environment grows into the thousands of calls. This section will examine several of the main issues that may affect how well call completion and call quality works.

Preparing the Head Office LAN to Support CallManager Clusters

One essential ingredient in CallManager clusters is the assurance of bandwidth between the servers themselves. Keeping in mind that the CallManager SQL server databases are the main data to be synchronized, larger head office installations might have some very large databases. The SQL servers can partially synchronize only changed data, but nonetheless this is critical data.

Therefore, many large VoIP installations employ the use of virtual LANs (VLANs) to the CallManagers so they can operate on their own dedicated bandwidth. One separate VLAN is used to carry data traffic, and yet one more VLAN is used to carry the voice IP phone traffic. This arrangement is shown in Figure 11.8.

Figure 11.8 Using VLANs to Assure Bandwidth to Devices

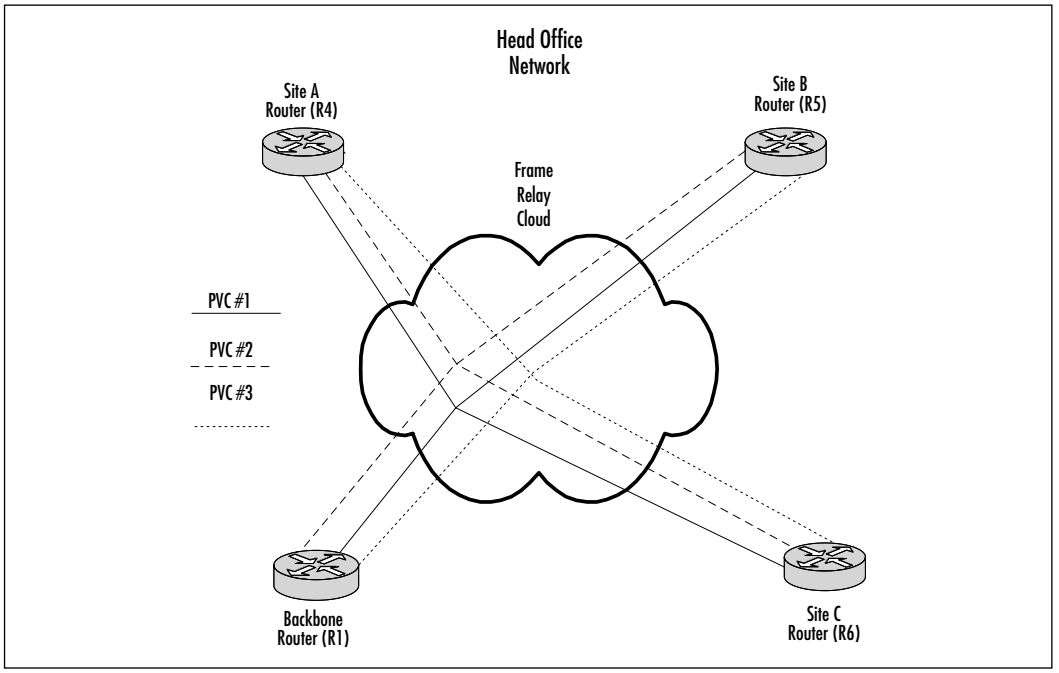


The VLANs between Site C and the head office backbone are just an example, because the real network would have the same VLANs extending to all sites, through all switches, and across the head office backbone to extend through

those switches as well. In this type of design, the routers will have QoS controls in place to assure that the 512 Kbps site circuits are properly managed and not clogged by any one process.

Instead of running VLANs across the network, one other choice when Frame Relay is used is to create multiple frame PVCs between the sites and the head office backbone router. This has the effect of creating logically independent networks across the same frame connection. While the Ethernet switches are segmented into discrete networks, the branch office routers do not propagate the site VLANs to the head office network but communicate to the head office via the PVCs. Figure 11.9 illustrates this LAN concept.

Figure 11.9 Using Site Segmentation with Multiple Frame PVCs



The advantage to this design is that traffic of one type can be directed down one PVC while other data types can get their own PVC pipe. The disadvantage is that this means much more expense than the VLAN methodology. For those reasons, the most accepted manner of WAN design for simplifying the LAN management is to use VLANs across the entire LAN and WAN topology.

Making Changes to the LAN to Handle Large Call Volumes

Before attempting this volume of traffic, we've found that the head office backbone must be up to speed as far as its rate of transmission and routing topology. The most modern LAN installations in the head office are using the Catalyst 6509 chassis with the Multilayer Switch Feature Card (MSFC) and Policy Feature Card (PFC) to enable Layer 2/3/4 traffic controls. The 6509 also employs the 8-port T-1 card with 24 DSP units on board this T-1 card. Lastly, the 6509 runs the 48-port 10/100 switch card that has in-line power ports.

With the LAN switches in place and operational, the baseline has been set to support CallManager clusters, multiple Exchange servers, Unity servers, and the VLANs between locations. The switched backbone network should be, at minimum, Fast Ethernet, but should also be Gigabit Ethernet between the servers and the 6509 chassis when possible and economically feasible.

The purpose for the T-1 DSP card is to provide conference calls, group bridges, and media mixing for AVVID applications that require such services. The T-1 card has 8 ports for PRI circuits, but you'll not be using the PRI for an actual circuit. Each of the DSPs can handle three mixed communications sessions at one time, so up to 24 conferences can be held at any given time. However, you'll need to reduce the 24 number by however many conference bridges you may dedicate to other compression protocols and dedicated functions, such as retaining one DSP for strictly internal office uses. Among these internal uses is the capability for mobile or home-based employees to call into the office and have dedicated processing capabilities.

DSPs are also used for transcoding purposes. Transcoding occurs when a device speaking one call type (such as an IP phone using g.711) contacts another device that uses a different call type (like an IP phone using g.729a) of compression. Since this is like two humans speaking different languages, the DSP acts like a translator to complete the call in an acceptable manner. Of these call types, conferencing only uses g.711 compression.

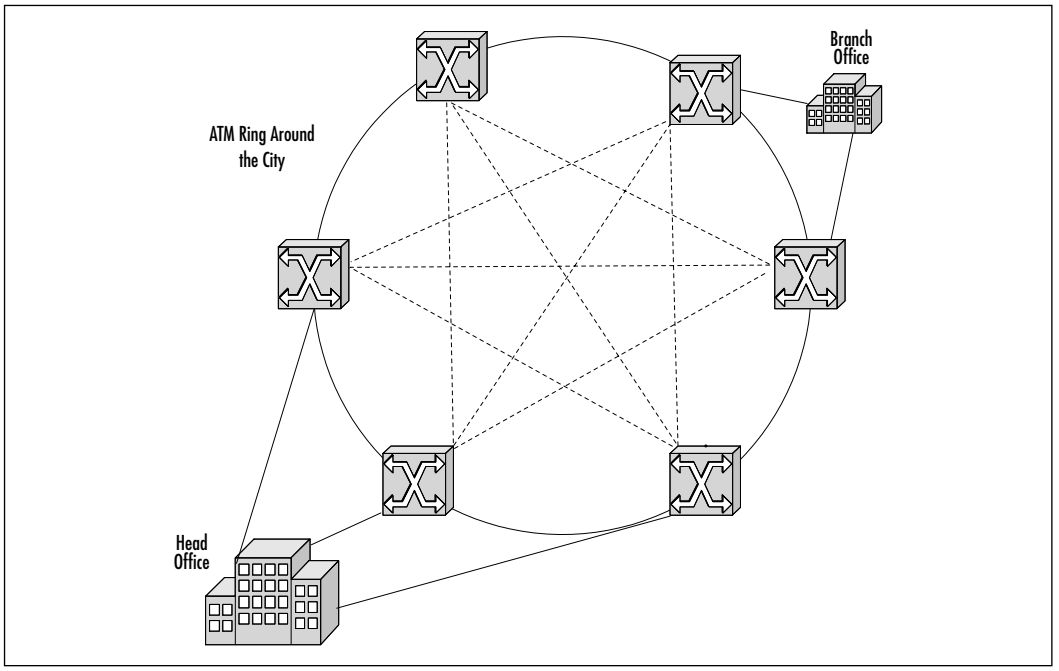
Providing Multiple Ingress/Egress Points to Sites

Providing a diversity of circuits for disaster prevention is one main reason for *not* configuring your network like the one in Figure 11.9. Just because a Frame Relay cloud exists in one major metropolis doesn't mean that there aren't multiple Frame Relay clouds extending all over the city. In Atlanta, Georgia, there are

14 major hub points circling the city to provide the backbone infrastructure. These 14 Asynchronous Transfer Mode (ATM) nodes criss-cross the city as well to provide a full mesh ATM network that is redundant, flexible, and has plenty of growth. The major providers of the backbone are a shared conglomerate of the carriers themselves, usually co-located in the existing Bell South central offices.

Because of this diversity, administrators can often get diverse routing of carrier solutions despite originating out of the same physical building. The circuits then go to different central offices, which connect to the different hubs across the city. Figure 11.10 shows an example of providing such route diversity.

Figure 11.10 An Example of Route Diversity



You can see in Figure 11.10 that the head office now has three possible points of access to the metropolitan WAN, and two possible points into the branch office. These multipoint facilities allow for emulated LAN protocols when ATM is used for both buildings. Also notice the building access to the ATM ring; those connecting lines are at actual locations on the building structure to permit diversified cable routing to and from the building structure. If a cable break were to occur, communications would continue no matter where the location of the cable break was, because the other circuits provide continuous access.

Designing the CallManager Centralized Solution

In this section, we'll discuss the centralized CallManager design. Each CallManager is capable of hosting 2500 clients, and CallManager can be clustered together in a logical manner to provide backup and redundancy to each organization. This section is devoted to presenting CallManager designs geared towards clustered solutions for a large enterprise. The reference to 2500 clients per server is a Cisco recommendation using the servers that they suggest. However, it may be more practical to use more servers for fewer clients per server. Some VoIP engineers recommend no more than 800 clients per server just to be safe.

Enterprise Dial Plans

Dial plans handle two types of calls: those within the enterprise, and those to outside users using PSTN services. CallManager has many types of configurations designed to handle these two, yet very extensible, configurations. Before talking more about dial plans, let's review the component parts of the dialing architecture, in the order of influence and control:

- **Route pattern** This is the layout of the number dialed that follows that country's numbering system.
- **Route list** The route pattern is interpreted and then sent to a route group, which is a group of devices that handle the actual call. Group 1 might have the best long distance rates, Group 2 the second best, and so on, so calls can be sent out the best possible (or least costly) gateway.
- **Route group** This is a collection of gateways, either H.323, PSTN, Skinny Protocol, or MGCP. Devices within the route group can order the delivery of calls in a preference list.
- **Devices** Skinny Protocol like the DT24, Catalyst 6000, and analog trunks; MGCP-based voice-capable gateway, the VG200; H.323-based gateways, all Cisco IOS routers; H.323-only devices such as the CallManager and NetMeeting endpoints.

The dial plan is the second most important topic within the VoIP environment next to a properly installed CallManager. The dial plan should be taken into very careful consideration and evaluated beside the current PBX solution. We'll refer back to Figure 11.1 as we design our dial plan. The three branch offices along with the head office environment will also host mobile users on the Cisco

IP SoftPhone, conferencing, and features such as call park and call pickup. So, let's assign a group of numbers:

- **Head Office** 6000 through 6999
- **Site A** 7000 through 7099
- **Site B** 7100 through 7199
- **Site C** 7200 through 7299
- **Conference Calls** 7990 through 7999
- **Call Services** 7980 through 7989

These numbers provide for sufficient growth for all sites in question, at least for the foreseeable future. Each of the four sites has their own local calling access, so calling overhead has been reduced but not quite eliminated. With these numbers, creating the initial route plans simply point to Site A if the dialed number has the last four digits of 7000 through 7099, while the others follow the preceding bullet points. The WAN is the first choice to find the destination, unless the gatekeeper says that there's not enough bandwidth to reach the destination. According to the dial plan, if the IP WAN isn't available or able to deliver the traffic, the route group sends the traffic across the PSTN.

Let's say that a call from the head office to Site A was attempted. The intended extension was 7005 and was called from phone 6105 by simply dialing "7005." The head office phone would then contact CallManager to place the call, which then contacts the gatekeeper to ensure that the desired amount of bandwidth exists for the call to reach Site A. Gatekeeper reports back to CallManager that the connection to Site A is not currently capable of supporting a g.711 64 Kbps call.

If you were at Site A, the area code is 703 and the prefix is 250-xxxx. Since the gatekeeper told CallManager that the WAN connection can not support the call, CallManager now looks to the route group for the next possible call routing mechanism: the PSTN. CallManager uses a function called *call transformations*, which handles the call by adding 91703250xxxx, where *xxxx* lets the 7005 be inserted into the dialing string. Thereafter, the call is routed out the PSTN circuit.

How does this happen? Two other functions are used by CallManager to choose the routing of a call: route partitions and the calling search space. Think of a route partition as an IP subnet. This is a distinct logical block that requires a router to send the IP packet to one place or the other. The calling search space is the equivalent to an access control list, which says where this partition can be routed to or from.

Yet another useful tool is called a *locations* definition. A location is just what its name implies—a region of calling devices that can be controlled and modified as desired. An example of this comes in controlling lobby and guest phones, from which long distance calls should not be placed by someone visiting in the lobby office (guests could call local numbers in the immediate city). These are two distinct locations, defined as "city-only" and "employees," where "city-only" could make just local calls whereas employees could call anywhere.

All of these are issues that arise when defining the dial plan, and should be designed and carefully thought out before any configuration tasks are completed in CallManager.

Installing Backup CallManagers for Redundancy

To ensure redundancy, at least two CallManager servers must be installed. The first one is the primary CallManager, which is used to make all changes to the users and the VoIP system in general. The second CallManager is the one that users will actually authenticate and have call control made through. The primary call manager should not be used for call control services. The reason is that call changes to the system are made on the primary server, which is reflected in the MS SQL Server on CallManager. This SQL Server then propagates the database changes pushed out to the secondary CallManagers on a regular basis. If users were to authenticate against the primary CallManager while changes were taking place, unpredictable results would occur.

Cisco claims the CallManager solution can support 2500 users per CallManager server. The reality is that this number of users will choke most infrastructures long before CallManager overloads itself, even with a primary and secondary CallManager. Such massive utilization is where a distributed CallManager solution comes into its best usage. However, it is possible a Gigabit Ethernet backbone serving the CallManager solution as a whole can virtually eliminate this bottleneck of Fast Ethernet. But if you upgrade the backbone, don't forget to upgrade the CallManager server to also support Gigabit Ethernet lest you simply re-create the bottleneck.

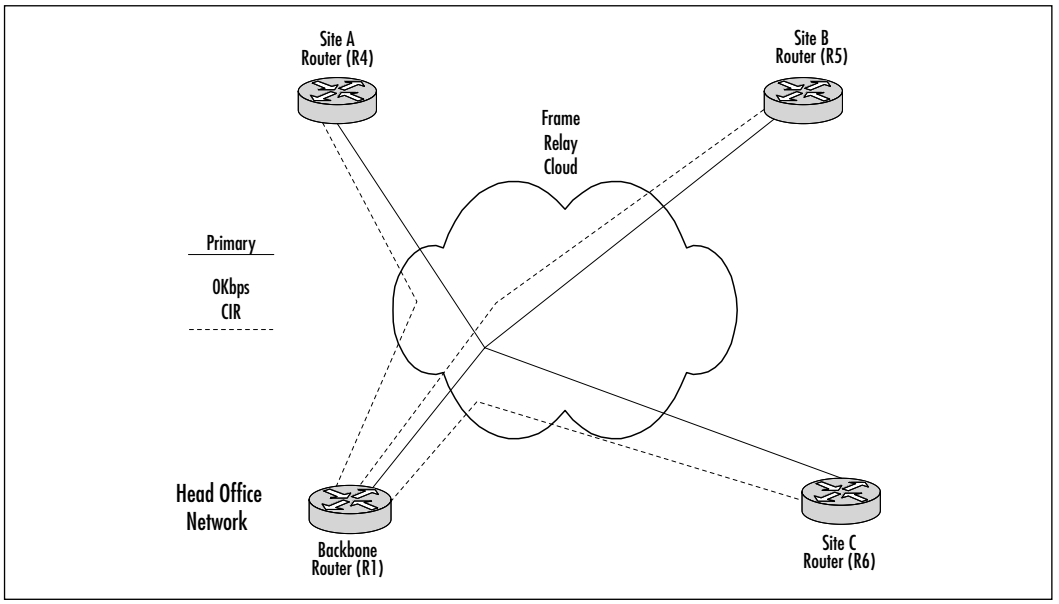
Assuring Constant User Connectivity to CallManager

To lose connectivity to CallManager would trigger intermittent loss of call ability, or cause the phone to lose its settings and possibly reboot the phone. This doesn't mean the phones will always have a TCP session to CallManager, but that by picking up the handset, it will result in a dial tone provided by CallManager.

Interesting solutions for call backups for the branch offices to the head office network include using dial backup provided by ISDN, or perhaps using a 0 Kbps CIR Frame Relay backup circuit to the head office network. If the primary site connection were to fail, then the routing protocols would detect this and know that the 0 Kbps CIR link was active and available.

Because this failure recovery is automatic and performed without the users' knowledge, it is almost a seamless recovery mechanism. It is only *almost* seamless because calls in progress would be lost since that call setup was performed across the original circuit. Figure 11.11 shows how this backup connectivity might look.

Figure 11.11 An Example of Route Recovery



By using routing protocols such as Enhanced Interior Gateway Routing Protocol (EIGRP) or Open Shortest Path First (OSPF), circuit downtime can be detected in seconds prompting the backup circuit to be activated just as quickly. But, if this backup solution uses technology such as ISDN Basic Rate Interface, then dialing time and Point-to-Point Protocol (PPP) instability might affect how well the routing protocol recovers the sessions and connections, so 0 Kbps Frame Relay might be the more stable choice.

Disaster Recovery for Centralized CallManager Solutions

CallManager is a set of hardware and software just like any other server solution, and it can fail just like any other solution. Therefore, we recommend a very good backup solution that can copy open files, handle SQL Server databases, and work with gigabytes of data. If the full enterprise CallManager solution is implemented, a 30 to 50 gigabyte tape backup is not out of the question. True, the initial solution is not likely to exceed 2 or 3 gigabytes for a few hundred users, but expect this to grow enormously when multiple CallManagers are deployed.

Designing & Planning…

Centralized VoIP: The Main Theme of Cost Savings

In our initial design in Figure 11.1, the solution is completely based on a centralized platform. If we were to estimate the user volume of a medium-sized corporation at 12,500 users at 10 sites, then the number of servers required to handle this is 6 CallManagers, 10 Unity servers, and 10 Microsoft Exchange servers for all voice mail needs. While this is indeed a huge investment in technology and in people to maintain it, your own decision to use VoIP must equally balance out the raw cost savings plus the inherent savings.

The direct cost savings is a bit of a masquerade at first. You must evaluate your current cost of local and long distance, PBX equipment capital and recurring costs, as well as recurring circuit costs. Over time, the organization should have a track record of the cost per user of a traditional PBX system, which will be essential to making the cost comparisons equal and honest.

Among those inherent savings is the ability to perform moves /adds/changes in the blink of an eye without having to wait for the PBX provider to come around and perform the tasks. With a moderately trained administrative force, the time to complete these changes can be reduced from days to minutes. The less obvious part of this is the ability to schedule and host conference calls with little to no notice at all, another benefit of the AVVID platform.

In the next section, you'll learn how and why you should either take an existing centralized CallManager and distribute it, or design a brand new distributed solution. No matter which direction you take, the principles are the same as that which you just learned, with a few twists.

IP Telephony Multisite Distributed Call Processing Solutions

In this section, we'll cover how and why an enterprise might want to distribute call processing technology. Among the reasons such a distributed design might be pursued is to prevent a loss of call processing should there be a catastrophic event on the head office networks where the centralized CallManagers and associated servers reside. Another reason for call distribution might include cheaper toll rates with regard to the branch offices' local telephone provider. This section will provide design information about these types of reasons for distributing the call processing.

CallManager Designs and Issues

To keep things in perspective, please refer to Figure 11.12 for this entire section on distributed call processing. The figure shows an example of how the processing servers might be distributed between sites.

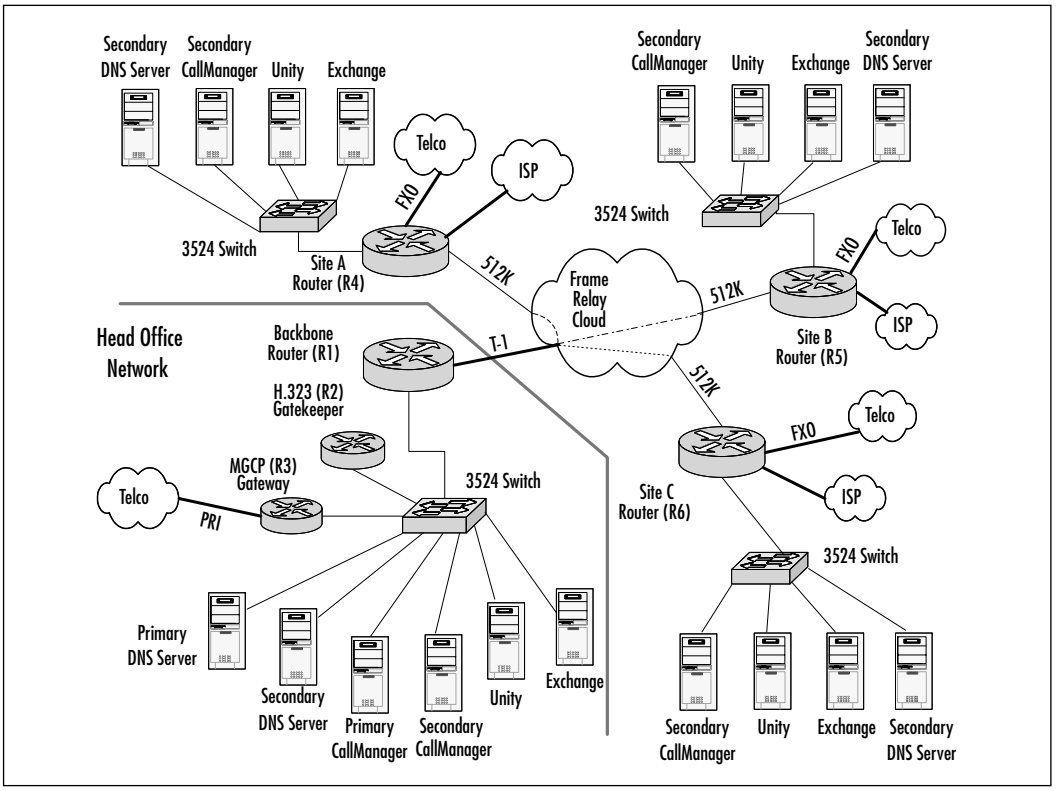
There are a few interesting changes to note here from our centralized designs:

- There are two CallManagers at the head office, one primary and one secondary.
- There are two Domain Name Server (DNS) servers at the head office, one primary and one secondary.
- The head office Exchange server is the mail bridgehead for all sites.
- Each site now has a secondary DNS server, a secondary CallManager, their branch Exchange mail server, and their own Unity voice mail server.
- Each branch office now has its own local Internet access point.
- Each branch office now has its own off-net access point to their local telephone provider.

These are important distinctions that will become more apparent as the design evolves throughout this section. The biggest issue is that each branch office has gained a large degree of autonomy and responsibility for maintenance of their

own systems, but this maintenance topic will be expounded upon significantly in the design phase.

Figure 11.12 Distributed CallManager Solution



Extending Enterprise Dial Plans to the Field CallManagers

The distributed CallManager environment raises all manner of concerns for the dial plan, and will mandate use of the locations, regions, and route patterns to discern the different sites. Because each site now has a CallManager, call search spaces and route patterns must be customized for each local environment. Gateways must be altered to reflect where the primary and backup CallManagers reside.

The centralized dial plan must be broken out to include the various local and mobile users. But currently, site-to-site calling is a function of invoking the IP WAN infrastructure and using up more of the IP WAN bandwidth. Since there will soon be more setup information streaming across the IP WAN, the IP WAN must be able to handle an increased level of traffic.

This means that to avoid excessive intersite calling, the dial plans on each site must be adjusted to use regions, device pools, locations, route patterns, and calling search spaces, and then each option must be configured to use these contexts. We previously defined phone numbers for each site, but now we must parcel out those numbers to the users. For instance, site A uses the numbers 7000 through 7099, which are divvied out between the office users and the mobile users. When changes are made to the sites' phones or call plans, the change is now made on the primary CallManager on the head office network, which then pushes out the changes to the site's local CallManager. There are site configurations that do not use this synchronization, so changes are then made to the local CallManager.

The CallManager setup will not have significant changes—what's important is where these changes are made. Also, to assure constant connectivity to the phones, you'll need to adjust each gateway to use the MGCP agent command and specify alternate CallManagers to use in the event the branch office CallManager goes down.

In the extended CallManager configuration for the servers, you can add remote CallManagers to your CallManager configuration, so you form a CallManager cluster. This causes field CallManagers to receive updates from the primary CallManager via the SQL Server replication of the databases.

Supporting Distributed Call Processing with Overall Design Changes

To move from a centralized to a distributed system, you'll have to consider a good many issues regarding CallManager to prevent a loss of functionality:

1. Review current WAN bandwidth utilization to ensure sufficient connectivity for the CallManagers. For each IP call on the site, allocate 64 Kbps for g.711 compression, and 20 Kbps for g.729a compression. This bandwidth requirement is to give consideration for concurrent call usage.
2. For each site CallManager, allocate 64 Kbps for synchronization between CallManagers.
3. For each IP SoftPhone, allocate 20 Kbps per phone, and use the low bandwidth CODEC for the SoftPhone users at the g.729a compression rate.
4. You'll be splitting out the CallManager functions on the head office network to the branch offices. You'll need to create the dial plans at the branch office CallManagers before removing the phones and dial plans

from the head office CallManager. Be sure you set up the calling search spaces and partitions that match the different sites.

5. Once you've accomplished Step 4, you'll need to add the field CallManagers to the head office primary CallManager, so that full dial plan synchronization can occur.

These are the most important tasks when creating the distributed dial plan out of a centralized design. Other incidental issues specific to your corporation will pop up, but these are the ones that will cause you to lose the most sleep. The best way to prepare for this migration is to ensure your existing centralized design is completely documented.

Disaster Recovery for Distributed CallManager Solutions

Distributed designs are inherently redundant, but not perfectly so. To ensure CallManager cluster communications and site access, some organizations create a second Frame Relay PVC between sites which connect the CallManagers together on their own network. This does not isolate the CallManagers from the local users, as each branch office router performs the intrasite routing function.

Configuring & Implementing…

Testing CallManager Redundancy

Having good backups are great and necessary, but this form of data security isn't enough. To ensure the CallManagers are properly configured, you should perform a live outage test to make certain phones find the redundant servers. The best-case test scenario is to shut down the branch office CallManager and ensure that calls still go through. To verify that the proper CallManager was used, use the **debug mgcp all** command in the field gateway with debugging turned on. A debug will show MGCP setup and communications between the gateway and the CallManager. Just be sure to do this from a console connection and not via Telnet, lest you overload the circuit with debugging information during the calls. From this debug information, the proper CallManager should be contacted for call completion.

Such redundancy does cost more, but you can be assured that the Call managers will have dedicated bandwidth between them without causing a loss of WAN performance. The only hitch is to ensure that all network routers and switches support VLANs between the sites. You'll also need local network administrators capable of handling the technical aspects of this design.

WAN Designs That Support Distributed CallManager

This section will make several changes to the WAN environment that will support our example distributed CallManager solution and assure proper connectivity between all sites. Keep in mind, these new facts and figures pertain to a solution for a medium- or large-sized corporation. Your organization may find some of these changes a bit on the expensive side.

There are really only two changes needed: to create a fully meshed and a partially meshed WAN architecture. The reasoning is simple. Either the site must have 100 percent WAN redundancy, or it can get by with a little bit of downtime. Either way, you'll find a mix of these two solutions viable for your needs.

Full Meshed WAN Designs

A *fully meshed* environment is one in which two or more WAN connections provide an ingress/egress point for each site. These connection points can use any of a number of technologies, such as leased line, Frame Relay, or dial-up ISDN. The point is to provide the WAN connection, regardless of the telco issues. Figure 11.13 shows an example of a fully meshed topology using Frame Relay.

Notice how each site is cross-connected to all the other sites? In previous WAN designs, only the branch offices were connected back to the head office router. But now, all the branch offices (or sites) have a connection to all the other branch offices regardless of which one processes what data. In this environment, you would experience the greatest demands in cost and network management, but you'd have to compare that burden to suffering the catastrophic losses of an outage that could cause a site to lose connectivity.

By having this design run a dynamic protocol such as OSPF or EIGRP, any link outage would be instantly noticed and traffic rerouting would occur just as quickly. However, if a call is currently in progress, that call would be lost but subsequent calls would be routed across the most available path. If you're interested in the cost of this solution, look at it from the standpoint that each site now has

five connections to the Frame Relay cloud. This consists of one of two possible connection types:

- One serial port on each router, dividing the T-1 capacity into chunks of equal capacity.
- Multiple serial ports on multiple routers to provide T-1 capacity to each of the five links leaving the site.

Figure 11.13 Fully Meshed Frame Relay Connections

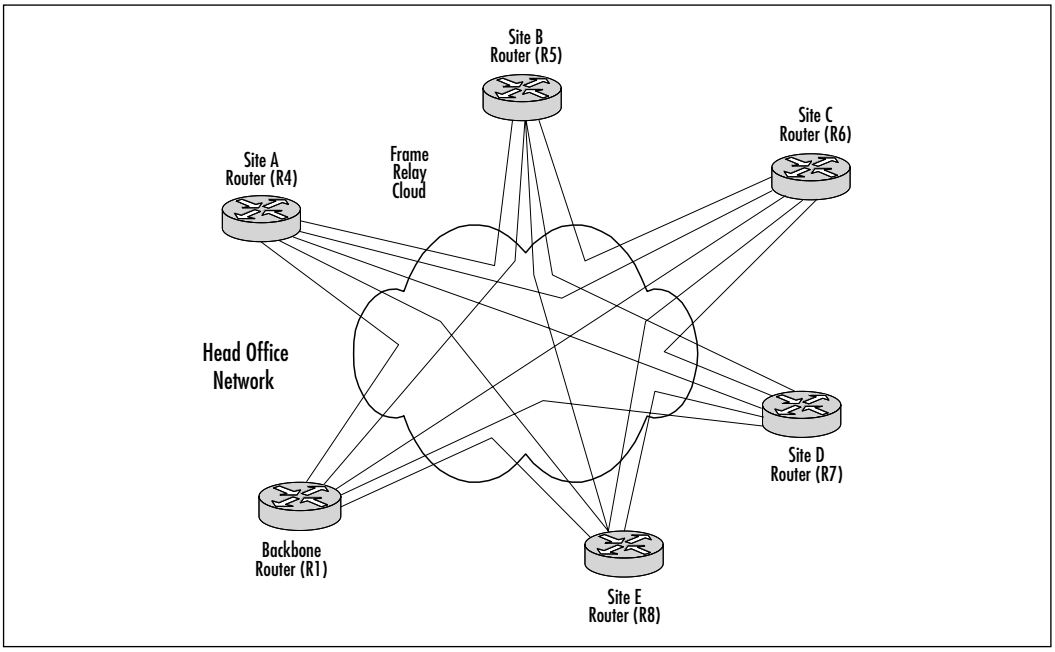


Table 11.1 shows an estimated capital cost of providing the 25 fully meshed WAN connections.

Table 11.1 Approximate Cost of a Fully Meshed WAN Design

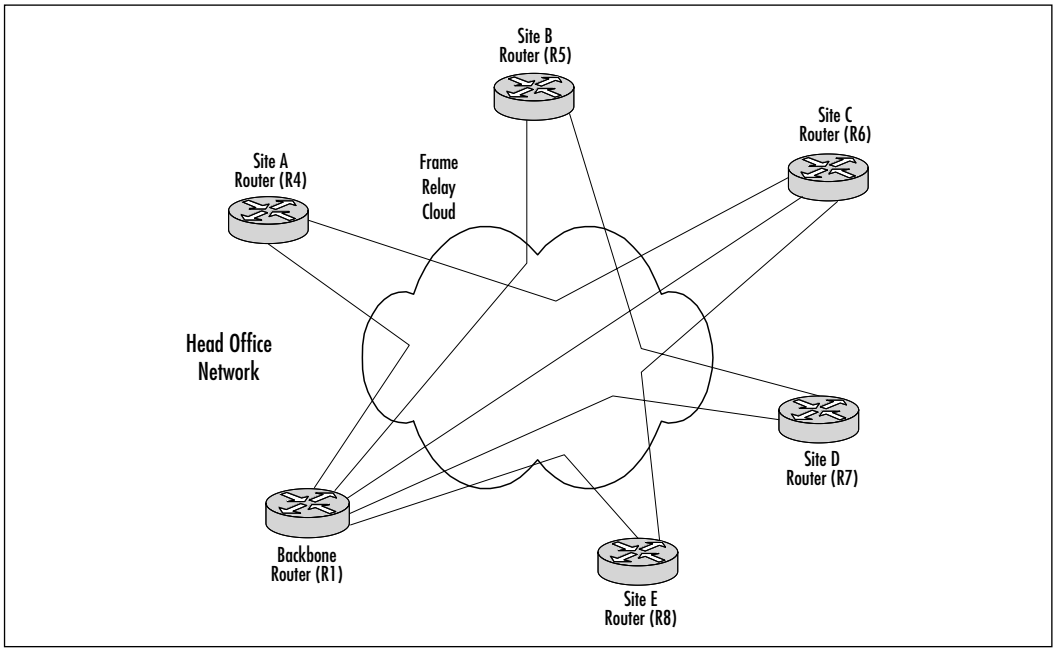
Item	Unit Cost	Quantity	Subtotal
3640 router	$9000	6	$54,000
NM-4T serial card	$1400	12	$16,800
T-1 CSU/DSU	$2000	25	$50,000
Total capital costs			$120,800

Not all areas of the country will have the same cost, but a T-1 frame connection might average $1200 a month in volume discounts, so this yields a monthly recurring cost of $30,000. These figures are purely estimates; your costs may vary.

Partially Meshed WAN Designs

In a partially meshed environment, like that in Figure 11.14, there are two connections to each branch office such that no one lost circuit will disrupt communications. Note that Site C has three connections by virtue of the Site E second connection.

Figure 11.14 Partially Meshed Frame Relay Connections



The fully meshed network of 25 frame PVCs has been reduced to 16 frame PVCs, at a significant savings in both capital costs and recurring monthly circuit cost. Table 11.2 shows an estimated capital cost of providing the partially meshed WAN connections.

Table 11.2 Approximate Cost of a Partially Meshed WAN Design

Item	Unit Cost	Quantity	Subtotal
3640 router	$9000	1	$9,000

Continued

Table 11.2 Continued

Item	Unit Cost	Quantity	Subtotal
NM-4T serial card	$1400	2	$2,800
T-1 CSU/DSU	$2000	5	$10,000
2621 router	$2500	5	$5,000
2621 WIC-CSU	$1200	11	$13,200
Total capital costs			$40,000

The monthly recurring cost now drops to a more manageable $19,200 a month, which is quite a savings. This design still has redundancy benefits, but cannot sustain more than 40 concurrent calls per site, giving ample room for data and CallManager traffic.

Determining Network Impact of Distributed CallManager Clusters

CallManager consists of the local databases, and of the SQL Server enterprise databases, which hold the entire configuration of the system. The information contained in the numerous databases is replicated to the secondary CallManager databases whenever changes are made to the primary CallManager. This section will take a look at the congestion issues that a CallManager cluster can cause in a fully distributed environment, how these issues can affect the enterprise, and what you can do to resolve these before they reach a crisis point.

LAN Issues for CallManager Clusters

The biggest concerns in using CallManager clusters are the security of the primary CallManager, how well the server is backed up and how well the SQL Server database is maintained. Regular maintenance should be run on the primary where the databases and log files are copied to another location for safekeeping. These files will grow over time in a large organization, so adequate amounts of disk space must be allocated.

Regular maintenance on the LAN side will cause bursty traffic, but not so much that a solid Fast Ethernet network couldn't handle it well. However, if the CallManagers reside on a subnet with other servers, such as database servers that transact e-commerce business, you might experience some difficulties. Another place that might cause a problem is the demilitarized zone (DMZ) area of a firewall, in which the CallManager ports may be blocked by rules.

If the opportunity arises, the servers should be placed on their own subnet away from other file, print, and database services. This might mean the allocation of a new router port, but it will guarantee that the servers not only have the proper LAN bandwidth, but that they have the proper security to prevent intruders and unauthorized visitors from reaching the CallManagers.

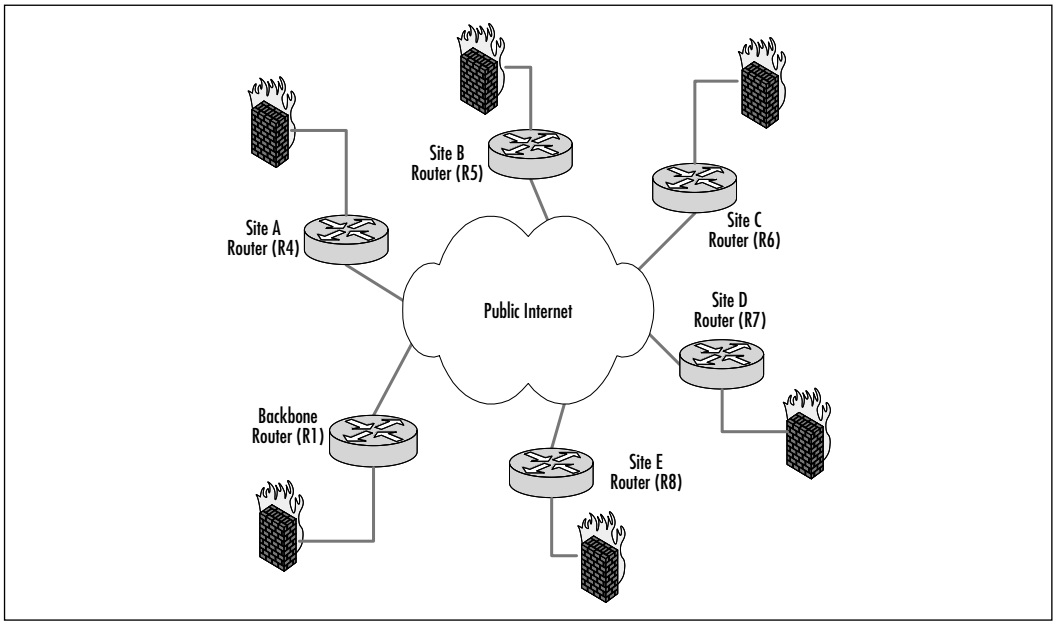
WAN Performance between CallManagers

We delved into WAN performance issues in a previous section, but have not yet discussed firewalling WAN access. This is because the Frame Relay connections provide inherent protection due to the private nature of the circuits, and don't require a full firewall. The Frame Relay routers do, however, have adequate access control lists for basic and extended security.

The last remaining WAN issue is the use of a virtual private network (VPN) to join remote and branch offices for this enterprise. A VPN is an encryption of the connection that joins two endpoints and causes the resources at either end to look like they are on the same network.

In this scenario, the traffic does not know it is flowing across the Internet, and people on the Internet cannot see the traffic that is encrypted. Figure 11.15 shows how this might be done.

Figure 11.15 Using a VPN for the Site Connections



In this figure, all sites connect to the public Internet using whatever connection they can find with adequate bandwidth for the solution. That's as far as the connection goes, however. Each firewall then forms a connection to the other site's firewall using an encryption *tunnel*. Once done, the two sites look and feel like they're on a local subnet—the difference is that the speed between the two sites might be a little slower because of the overhead the encryption engenders upon the Internet connection. Next, to form a connection to the branch offices, the head office firewall creates the same encryption tunnel to the other branch offices, thus creating the head office network while running across the public Internet.

This arrangement works fine, but it strictly depends upon the stability of the public Internet for its usefulness. Under these circumstances, the CallManagers and/or the CallManager cluster may suffer if the circuits from the site to the Internet are slower than a full T-1. Practical experience suggests a limit of about 25 users per site.

Unity Messaging Issues

Because Unity is based upon Microsoft Exchange Server, any issues with voice messaging will coexist with LAN and WAN connectivity issues. A typical 60-second voice message will generate a WAV file in Unity e-mail around 400KB. Longer messages can exceed 2MB, so slower links at any part of the solution will be further aggravated by the size of the files. Unity synchronizes with the main Exchange mail server, so there must be reasonable bandwidth available for this to occur.

Designing & Planning…

Distributed VoIP Spreads the Workload

The advantage of the distributed solution is that it tends to keep the processing local to the source calls. This isn't a hard and fast rule since many businesses make lots of long distance calls, which tend to leave the site. If the main intracompany calls are back to the head office, the calling decision is still made at the branch office's CallManager.

There are advantages to having multiple gatekeepers, where each one is located at one of the branch offices. This is an interesting configuration, and is useful when two or more branch offices are geographically

Continued

closer together than either one is to the head office. This is especially true if and when more calls are made between branch offices than are made to the head office or outside destinations.

At any rate, the distributed environment normally removes a measurable amount of loading from the head office network, but this must be done carefully and planned out very well. If not, then the rising capital costs can drive the overall solution to an unmanageable level of expense, one which has virtually no return on investment.

Multisite AVVID Solutions

This last major section of the chapter will be devoted to the remaining major functions of the AVVID family: IP Television (Cisco's IP/TV), IP Video Conferencing (Cisco IP/VC), and the associated family of tools for the AVVID family. In order to properly implement these solutions, the underlying network must be capable of correctly supporting multicasting across the switches and routers. Before discussing the AVVID applications themselves, a discussion of multicasting should give you the basic knowledge necessary for understanding each application.

Designing the Enterprise IP Network for Multicasting

Before discussing the applications, let's get into some of the LAN and WAN issues you'll encounter for this larger scope of multicast applications. Cisco Group Management Protocol (CGMP) is used by the routers to register and receive multicast updates from various sources, and to hold status data on the multicast environment. The CGMP family of commands is used to control how routers and CGMP-capable switches join groups, request IGMP reports of which devices register for multicast, and use similar multicast controls.

Multicasting is used on the reserved Class D address range of 224.0.0.0 through 239.255.255.255, and has several reserved areas. The multicast range of 224.0.0.0 through 224.0.0.255 is reserved for routing protocols to share updates between other routers. This range will not (or should not) ever be forwarded across routers; rather, this address range is used in-between routers for these updates. Layer 2 multicasting addresses use the Media Access Control (MAC) layer to control the data; starting at 01:00:5E, the total range is 01:00:5e:00:00:00 through 01:00:5e:7f:ff:ff as the lower bit range.

This leaves the upper range of MAC addresses free in the multicast range. When a device in a switched environment wants to join in a multicast, this

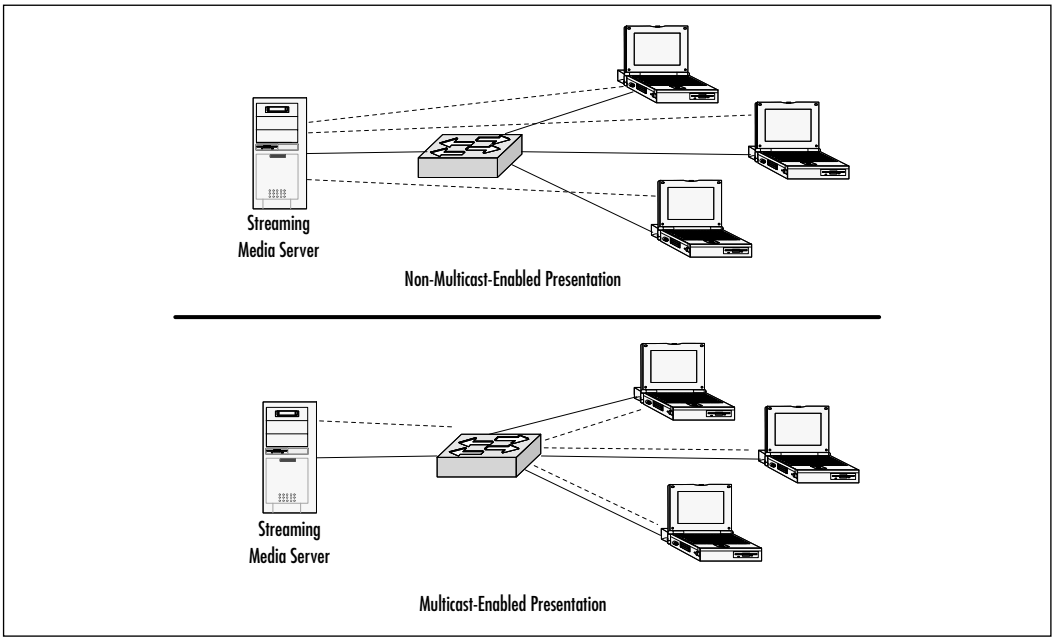
device signals the switch via CGMP and its MAC address is used to logically AND with the group address, resulting in the new registration MAC address.

This causes the end device to register with the switch as being a member of that multicast group. When the sending device begins the stream, the stream reaches the switch where it is then sent out the switch port to the end device, and not all the other devices connected to the switch.

The next two sections of the chapter are devoted to the only places where multicasting will be an issue, which is across the LAN and WAN connections. We'll first examine how the local head office and branch office LAN routers must be configured and administered to support this multicast environment.

Before you can get into this discussion, you first must understand exactly how multicasting works. Figure 11.16 illustrates the use of multicasting in the switched LAN, and shows a typical connection between the streaming media server and three client workstations. There are three distinct communication sessions going on, which is then three times the individual data rates. In other words, a single streaming media presentation that might use 6 Mbps of bandwidth for one user would now consume 18 Mbps. By contrast, multicasting ensures that only one stream of data is sent between multicast network devices, and those devices then provide the individual streams out to the recipients.

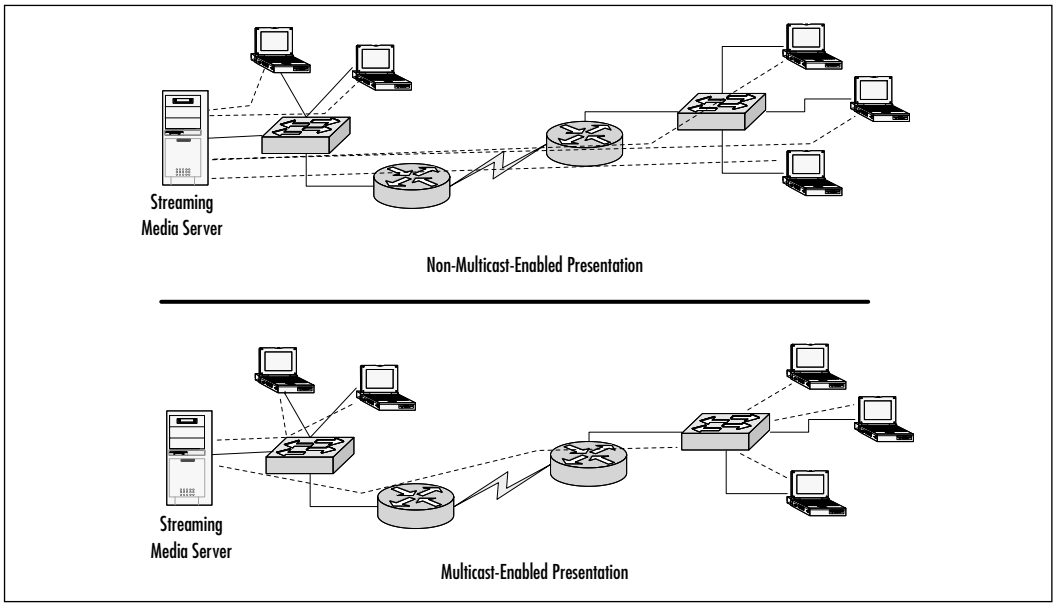
Figure 11.16 Examples of the Use of Multicasting in the Switched LAN



This figure shows that multicasting on the local switched enterprise serves to reduce the volume of traffic on that enterprise, and requires the use of enterprise edge and core switching solutions such as the Catalyst 4000 and 6000 Series core switches and 2900 and 2800 edge switches. These devices must be running the enterprise CatOS services to exploit these controls.

In Figure 11.17, an example of using multicast controls in the WAN clearly has its benefits along with a few bits of the WAN protocol overhead.

Figure 11.17 Examples of the Use of Multicasting in Routed Enterprises



There are numerous controls used to determine how multicast traffic makes it across the enterprise. Our example only uses two routers between locations, but exemplifies the functionality of multicast controls in the routed backbone.

Configuring the Routers to Support Multicasting

In the routed network, multicasting takes on a new form of traffic management. Multicast distribution trees control how and where the data traverses the network to ensure those devices that want the data receive it, and all others never see the traffic. This is essential to accomplishing what multicasting does—reducing network overhead while reaching out to multiple users.

In the Figure 11.17 example, you saw that using multicast controls effectively used only two streams of data to reach the simple network configuration. This is

due to the usage of what Cisco calls a rendezvous point (RP) of distribution. This designated router becomes the root level router for the purpose of multicast traffic. When the source device begins sending the multicast traffic, the traffic must first go to the RP, which makes the determination regarding which users will receive the multicast.

Another functional control of multicasting is sometimes called Reverse Path Forwarding (RPF). A router's job is to examine the destination address of the IP packet, and make a determination of how to get the packet to that location by forwarding it to the next most logical router. The decision-making router does not need to know the source address, just where to deliver it. RPF, on the other hand, requires that the decision-making router know who the sender is, so that the multicast traffic can never be sent back down the same path from which it originally came. This action ensures that the multicast traffic will never get into an endless loop and bog down the network.

Another protocol that Cisco uses is called Protocol Independent Multicasting (PIM). PIM gets its name from the fact that it actually uses the unicast lookup tables in the router for part of the decision-making, and PIM makes the RPF decision. Collectively, these two functions ensure that the multicast traffic is delivered regardless of whether the routing protocol is OSPF, Border Gateway Protocol (BGP), or EIGRP.

PIM has two modes of operation. *Dense mode* makes certain that multicast does a brute force delivery of the multicast all over the network, and builds the distribution tree upon the delivery through the routed architecture. Dense mode doesn't use the concept of a rendezvous point, since this is a forced distribution. Dense mode is not particularly efficient, which is why sparse mode PIM exists. *Sparse mode* actually works on a pull model in that only the registered and desired end users designated for multicast traffic actually receive the traffic, which is more like the distribution tree method previously mentioned.

Cisco has an implementation of a mixed mode, called *sparse-dense*, to address the best points of both modes. If the multicast data reached an area where all connected users wanted to receive the multicast, dense mode would be the better choice and it would use less overhead on the routers and switches—however, you'd not want to use dense mode throughout the network. Sparse mode accomplishes the latter part better than dense mode. In this case, sparse-dense would determine that sparse be used to get the multicast to the RP, and then dense mode be used to blast the data down the final leg to the users.

In PIM sparse and dense modes, the sender and receivers must enter their desire to be in the multicast group by registering with the RP nearest to them.

One problem might exist in a very large enterprise, however. These registrations take place in a per-domain environment for the multicast community. What if this large enterprise had several multicast sessions running? It is possible that multiple receivers might want to see two or more sessions on their PC at a time.

Multicast source discovery protocol solves this issue by allowing the RPs to periodically share registration information for the user closest to the RP, or that might be from different multicast domains. The RP of the first domain sends a discovery to the RP of the other domains so this RP can learn what other multicast services the RPs cover.

Wide Area Network Considerations

The considerations for the wide area is nothing more than placing proper traffic controls at the ingress and egress points. Multicast controls for the WAN specify much of the same type of multicast controls as the LAN side, but provide the ability to add multicast routing protocol controls and compression to the data stream. These WAN controls ensure that not only is the multicast traffic properly handled, but the proper form of WAN compression and queue controls are put into place.

VPN connectivity was discussed earlier in this chapter as one mode of communications, but it is not suitable for multicast applications. The reason is that all devices participating in the multicast sessions must be multicast-aware and also be running the proper software for multicasting. From the end user to the backbone routers between users, devices must have the proper IOS firmware that is identified as multicast capable.

Also, when using either of the IP/TV and IP/VC solutions, the data streams will be pushing significant amounts of data.

The use of field routers beginning with the Cisco 2600 series at the branch offices then becomes an absolute necessity so that the proper amount of flash and shared memory can be installed in the routers. The same is true of the Ethernet switches, given that speed alone is no longer the defining factor to ensure the proper performance of the solution.

Finally, some administrators might even consider running a parallel WAN circuit between sites that run only the multicasting traffic.

This is just a second WAN circuit between the affected sites, but routing controls are used along with route maps to direct the multicast traffic to the destination. Such circuits are dedicated for this usage, and offer the benefit of not conflicting with the existing VoIP traffic or other network connectivity.

However, there's the cost factor again rearing its ugly head. Using Frame Relay is a good option with a 0 Kbps CIR, burstable up to the full T-1 capacity. This means that you'd pay a minimum amount just to have the circuit (something like $400 a month to put it in place, then as much as $1500 a month when it's used). But what if you only use it an hour a week?

Some circuit providers will offer the $400 rate for access, then charge you on a per megabyte basis for actual usage. Some providers charge as little as 50 cents per megabyte; others as much as $2 per megabyte. You'll just need to be very careful about the type of contract offered, and the limitations it may impose.

Cisco's IP Television Solution

In Chapter 10, you saw how IP/TV is used to reduce the cost of television presentations. While many people do not think TV is a productive part of head office life, there are many excellent uses for enterprise-wide IP/TV solutions, both productive to the company and the employees.

1. Many academic institutions provide taped lectures by their instructors for the adult student population who typically work during the day, or cannot travel to the class during the week. These students could receive their academic lessons adhoc to the classroom experience.
2. Employees could see ad-hoc meetings they missed, but that were recorded for posterity. This includes Q&A sessions conducted during the meetings.
3. Do you have sports fans in your organization? Perhaps you'd like to offer something for employees to watch during their lunch break? Try broadcasting the Olympics over your network via the CNN news feed!
4. If you had prospective new customers you wanted to provide services to, why not show them what your company can do via a head office capabilities presentation?
5. Replay the CEO's annual company-wide presentation when new employees come to the company.

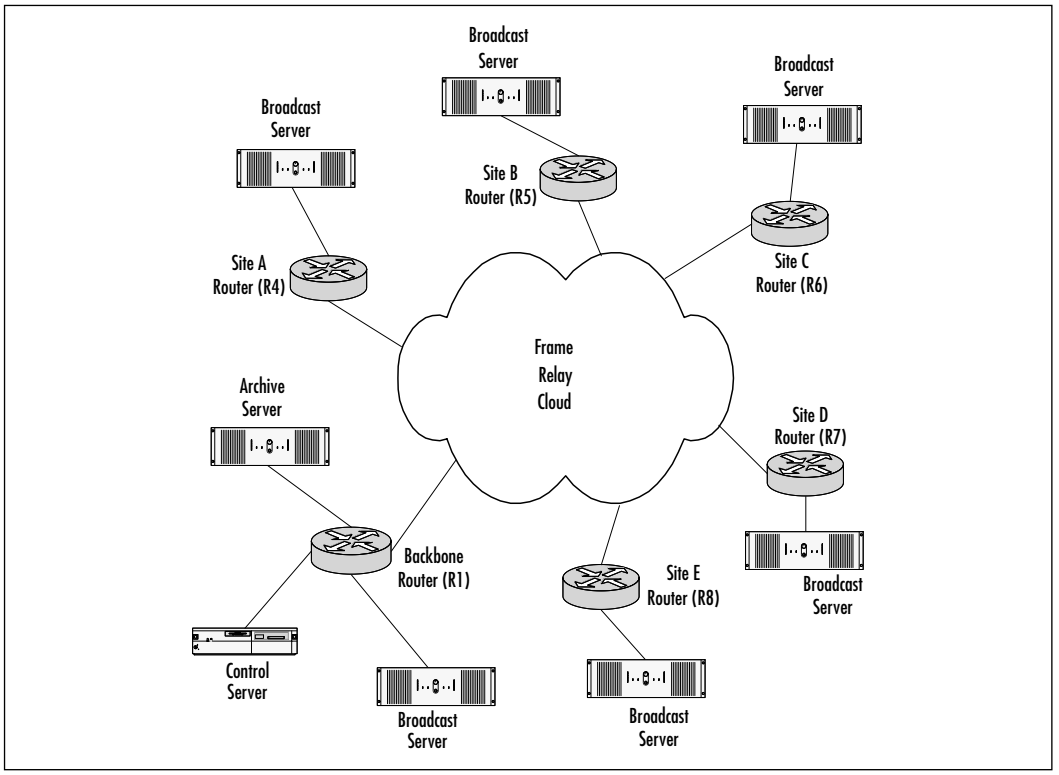
Using IP/TV with Branch Offices

We've covered all the LAN and WAN issues that affect both Voice over IP and multicast applications. So, from a theoretical standpoint, your LAN and WAN are ready to handle any AVVID solution you can throw at them. IP/TV takes a constant 4 Mbps of Ethernet (uncompressed) connectivity to get to each user's

desktop in full motion video. The broadcast server is a Cisco hardware platform that sends out the traffic to all subscribed users.

The proper design solution to assure continued data connectivity to the branch offices is a full T-1, of which 768 Kbps can be reserved for the compressed data stream to the site. If the WAN links are slow, consider putting a broadcast server out at the branch office. One Control Server can service quite a few broadcast servers, so this is no immediate issue unless you have a few dozen broadcast servers. Figure 11.18 shows a typical multiple site design for Cisco's IP/TV solution.

Figure 11.18 Cisco's IP/TV Solution in a Multiple Site Environment



This distributed solution is more appropriate since the site's broadcast server can receive presentations late at night when the main control server distributes the presentations. In this method, the broadcast servers for multiple sites can receive a high volume of traffic late at night when no one is using the circuit, making sure the presentation is ready the following day.

Choosing Devices for Enterprise IP/TV Solutions

The IP/TV Control Server device is the registration and policy control for the system. It handles coordination of the broadcast, the encoding of the source (such as from a video camera), and controls information for the IP/TV client software installed on the desktop computer. This server is capable of assuring that only the desired clients receive the broadcast, and can encrypt the broadcast if need be. The IP/TV Broadcast Server handles the actual processing of the streaming media, and distributes it to the destination based upon the control information hosted in the Control Server. Each Broadcast Server processes only one stream format per session. If you needed to present in MPEG-1 and MPEG-2 formats simultaneously, you'd need one Broadcast Server for each media type. The Broadcast Server accepts source input from many locations, including those from live feeds, VCR tapes, DVDs, and live recordings in local studios. The Archive Server handles broadcasts that must be played back as an ad-hoc presentation. A one-hour full video presentation in MPEG-2 format can easily consume 45MB of storage space, so this server should have suitable storage capacity and have the flexibility to allow for growth. You could also use a Storage Area Network (SAN) to house these source files. The IP/TV Client Viewer is the software application that runs on each client computer which receives the broadcast, and effectively completes the multicast registration to the Control Server. This registration uses the MAC information on the network adapter as the method of assuring proper delivery.

Cisco's IP Videoconferencing Solution

Cisco's IP videoconferencing (IP/VC) solution requires the use of a number of devices (see Chapter 10 for more details). Cisco solutions are backward-compatible with the older H.320 standard to help customers migrate to the latest standard without a loss of investment. The solutions use the following devices, although not all are used at each deployment:

- **Multimedia Conference Manager** This is an IOS-based gatekeeper solution that provides call admission control, proxy management, call detail records, and security controls for each VC connection.
- **IP/VC 3510 Multipoint Control Unit** Serving as many as 15 connection points, this device provides gatekeeper controls for small sites of speeds up to 768 Kbps.

- **IP/VC 3520/3525 Video Gateways** These two units provide BRI and low speed (384 Kbps) connection video processing of H.320 to H.323 schemas. The 3520 supports BRI capacity while the 3525 supports a single ISDN-PRI connection to legacy systems. Both units provide internal gatekeeper functionality.
- **IP/VC 3540 Multipoint Control Unit** This unit is much the same as the 3510 unit, but it supports up to 100 connection points and the controls and increased processing capacity needed for this higher number of connections. It also supports connection speeds up to full T-1 capacity.
- **IP/VC 3540 Application Server** This Windows NT server is used to provide applications level support for whiteboarding and application sharing.
- **IP/VC 3540 Gateway Module** This piece of hardware provides H.320 to H.323 conversion up to 384 Kbps connection speeds with the most popular video types and formats.

More than one of these devices might be needed at each of the branch offices, depending on what it is you're trying to achieve. The Chapter 10 discussion largely holds true here as well, but now there are many more sites to deal with. One such example can be seen in the academic world, where whiteboards are often used by instructors. Branch sites that might only have ISDN capability would use a 3520 gateway and perhaps one 3540 application server for each class at the branch location. Other sites that have ISDN PRI capability would tend to use the 3525 gateway plus the application server for its classes.

Using IP/VC for Multiple Sites

Using IP/VC provides enterprises with a means of face-to-face communication without the expense of travel costs and time. With the advent of personal satellite communications, it is now possible to have 400 Kbps available to you no matter whether your location is on a mountaintop or in a desert.

IT engineers can use IP/VC as a network troubleshooting tool. Using the little net camera and a voice sound card, you can be at opposite ends of a campus and still have a chat about troubleshooting problems and issues. Assuming you're not trying to fix the very section that IP/VC depends upon, this ad-hoc conferencing is great for showing someone on a piece of paper what was found, or for conducting whiteboard sessions to work out the problem.

Why IP/VC Can Be Damaging to an Enterprise

While Cisco's IP/VC product is good, the implementation of it in a network that is not fully prepared for conferencing and/or multicasting can leave a devastating impression upon the users and the purchasers alike.

The worst thing that installers of Cisco's IP/VC solution can do is assume that the network is stable just because the customer told them it is. The installer should perform a variety of tests to assure proper connectivity, sufficient bandwidth in the LAN and WAN alike, then set up only a test bed for a short trial run.

This sounds like a highly cautious approach, but it's done for a very good reason. The deployment of a multicasting application onto a network not ready for that application is destined to fail. You'll spend hours or days troubleshooting video and audio problems only to find that the network was never stable in the first place. There may be cabling issues, incompatible network devices, or simply just not enough available bandwidth to support the application. This is nothing that a careful plan and a good network analysis won't solve for you.

Lastly, plan your bandwidth needs carefully. You've seen in Figures 11.16 and 11.17 how the switched and routed environments are affected by multicast traffic. IP/VC is a little more complex than just pure multicast in that IP/VC is two-way traffic of both video and audio, which might not be a multicast stream if there are many ad-hoc conferences placed directly between site users. Remember that users between sites themselves can place ad-hoc conferences without going to the head office site, yet the conference must be controlled and authorized by the gatekeeper on the head office site.

Uncontrolled, these ad-hoc conferences can flood the various parts of the network and prevent scheduled conferences from ever taking place. It is these types of user-initiated applications of the overall AVVID solution that can cause harm to the enterprise.

Creating the Auto Attendant

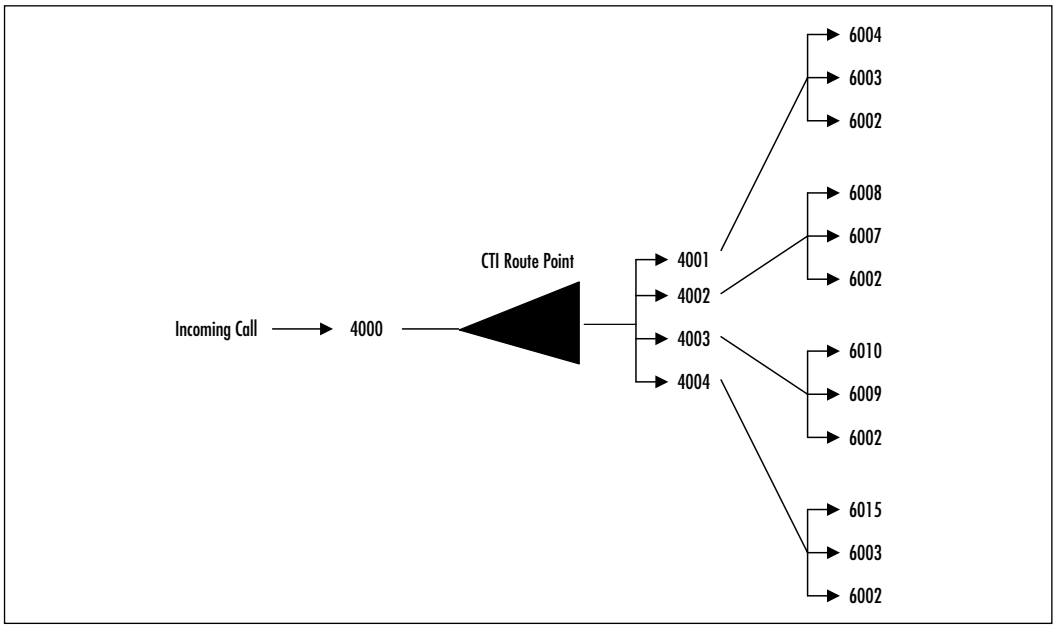
Cisco Auto Attendant is the tool that receives incoming calls and automatically routes the call to a pool of prospective recipients. The tool functions by using a Cisco Telephony Integration (CTI) point as the first instance of reference. Next, each of the CTI points are mapped to a series of user phones such that a hunt group is defined. Figure 11.19 shows an example of how Auto Attendant works.

Notice that the route numbers are sequential, which is a requirement, but that the end user numbers are not sequential. Also, take notice that the user number 6002 is at the end of each hunt group of numbers. If no one in that group

answers, 6002 will always receive the call. The steps needed to create the Auto Attendant are:

1. Create one CTI route point.
2. Create four sequentially numbered CTI ports (we used 4001 through 4004 in our example).
3. Create the user for the IP AA application. It is best to employ a new user and not an existing one, just to keep the services separated from standard users.
4. Install and configure the Java Telephony API tools on each Cisco CallManager that will use the route point.
5. Configure the JTAPI application to use the route points established in Step 2.

Figure 11.19 Auto Attendant CTI Call Routing



Auto Attendant is a nice tool for automating incoming calls to groups of users, such as call centers and service desks where there are multiple end users. While Auto Attendant is not a clustered solution itself, if you cluster CallManagers, then the CTI route points will be accessible at those respective sites, and at any other sites permitted to reach the route point in their calling search space.

Using Web Attendant

Cisco Web Attendant is an exciting application that replaces the traditional switchboard with a software-based PC tool set for processing incoming calls, forwarding calls, and performing other related tasks. Web Attendant can be installed on any Windows 9*x* client computer, Windows NT4 workstation, and Windows 2000 in the future. Web Attendant is associated with one phone, such as the front office secretary.

Web Attendant integrates into the Light Weight Directory Access Protocol (LDAP) for user information and services. LDAP is installed and running on the Unity server and on the Microsoft Exchange mail server, which is where Web Attendant pulls its information. You can have many Web Attendants on the enterprise, such as one at each of the locations where the secondary CallManager resides. This is a very good idea, as Web Attendants can back up one another when the CallManager cluster is established.

The really cool part about Web Attendant is that being Web-based, you can see the state of the IP phones of those users that Web Attendant covers. The phones show up as an icon within Web Attendant, showing if the phone user is online, in a conference call, or available to take a call. This means that the attendant doesn't have to call the user first to determine availability, because Web Attendant shows the availability.

IP Interactive Voice Response System

The Cisco IP Interactive Voice Response (IVR) system is a Customer Call Center tool used to automate tasks for providing information feedback to callers. If you've ever called your telephone provider for information about your account, you were asked to press 1 for billing inquiries, 2 for changes to your service, 3 for the collections department, and so on. When you press a button, the IVR runs a script that now sends you to another IVR.

Let's say you pressed 2 to make changes to your service. The initial IVR ran the script that took you to that set of services, which might be 1 to add, 2 to delete, and 3 to modify an existing service. If you pressed 2 to perform a deletion, you might be asked to type in the phone number to be deleted, at which time you're likely to be prompted to press 1 if this is the correct number, or 2 if it is incorrect. All the while, you've using a very large IVR system that might have as many as 100 or 200 IVR systems at this one call center. However, your usage for an IVR within an enterprise might be to offer automated services to the internal users, or to your customer base on a private IVR network.

The IVR devices can be located anywhere on the enterprise network where IP routing is configured. You might have several Customer Response Applications (CRA) engines located across the branch offices, placing the functionality closest to where the work flow occurs. Also, you might place the CRA engine at the branch office where the CallManager is located, since the IVR function integrates into CallManager as either a pure software function, or a hardware IVR integrated into the CallManager software control. The CRA engine uses the LDAP configuration for the users to determine startup execution and configuration.

Designing & Planning…

How Long before AVVID Really Matures?

AVVID is here, it works, and is deployable—the switching and routing infrastructure readily supports AVVID now. The biggest drawback at the moment is not so much the cost of the AVVID solutions themselves, but the changes to the infrastructure required to support the routing and multicast controls that allow AVVID to work flawlessly.

The second major factor holding back AVVID from full maturity is the need to run a full Cisco infrastructure solution. Many enterprises are a mix of old world electronics, or are die-hard Nortel Networks or 3Com solutions. While Nortel Networks does have a Voice over IP solution, they don't support the AVVID platform. So, if you want all of the AVVID solutions along with all the benefits AVVID brings to the table, you'll have to migrate to a Cisco platform. And that is the crux of the problem.

After performing many network analyses for migrations to Cisco, and creating project plans for the migration, this author has found that the immediate cost far outweighs the long-term benefits. Too many infrastructures require new wiring: routers (even if Cisco) may be using very old operating systems, the Ethernet switches can be old 10Mbps systems and may not have the CatOS to support multicasting, and the WAN links may be ancient 56Kbps DS-0 circuits that would require new CSU/DSU units everywhere.

So, is it a Cisco problem? Not really, because your infrastructure may never have kept up with the newer generation networking products, not even the ones that were one or two generations behind the most current. This means that to purchase the latest product supporting

Continued

AVVID, let's say a router, the router to be replaced is two or more generations old. Cisco does have a trade-up program, but only for devices that are usually only one generation back. This means that you would probably lose the value of the older router, even if it were working perfectly fine in the old environment.

However, there are good indications that Cisco is bringing their network hardware costs more inline with other vendors, which creates an excellent avenue for customers of older networks to upgrade. Even in hard economic times, Cisco devices can be purchased and networks upgraded in stages to not only modernize an enterprise, but also purchase the new devices as voice-ready and multicast-capable units. These AVVID features just aren't used at the time of the upgrade, but can be easily acquired as the need for the feature warrants.

This is the essential problem in any network. Sometimes not keeping up with the Jones's can be a very costly situation when new applications come around.

To generate the IVR flows, you'll use the CRA editor to create and manage the scripts that define the flow of action when the button is pressed. The editor works under Windows 9*x*, Windows NT v4, or Windows NT/2000 client computers, and sports a drag-and-drop feature to create many of the flow options and definitions. These flows are then stored on the CRA engine for later execution. If something goes wrong with the flow, the CRA engine has a good debugger for watching and fixing the flow execution.

There are three options for installing the IP IVR functionality:

- A pure software-based function installed on the CallManager, which provides 2 IVR ports and 8 auto attendant ports, or all 10 ports configured as auto attendants.
- A dedicated IP IVR solution running on a Cisco MCS 7825 server supporting up to 30 IVR ports. The default package ships with a 4-port license.
- A dedicated IP IVR solution running on a Cisco MCS 7835 server that supports from 24 to 48 IVR ports with a license for 24 ports shipped with the default package.

If you need more than either of these products offer, then you'll have to install additional servers or packages as might be required.

Summary

Head office environments always place high demands upon data processing needs, and the Voice over IP solutions along with the remaining AVVID solutions increases those demands quite a bit. You've seen in this chapter how centralized and distributed solutions work, what requirements they place upon the enterprise, and what risks and costs are exposed to the enterprise. You've also seen how various design changes might be required for the LAN and WAN to support multicasting requirements. Differences between CallManager implementations were presented so that you have a good idea of what changes are involved when considering centralized versus distributed solutions.

VoIP solutions can bring big benefits to a corporation provided the existing network infrastructure is capable of supporting high-speed data operations. The corporation's WAN must be extensible for the increased bandwidth needs that AVVID requires, yet be flexible enough with its hardware solutions to not bankrupt the company.

You've been exposed to the same AVVID solutions we deployed in the single site solution, but now these tools are being used in multisite applications. The amount of servers grew, the total site bandwidth grew, and the administrative requirements grew. Basically, everything just got a little more expensive and harder to deploy.

We've presented enterprise solutions for multiple sites for Web Attendant, Auto Attendant, and the Cisco IP IVR solution that might be beneficial to the enterprise. These tools are not always deployed in every AVVID environment, but can make use of the multicasting controls instituted for the VoIP and CallManager systems.

You've seen how we designed these solutions, how to modify them at a high level, and then what to expect in the way of systems management for daily and emergency recovery operations. When you make the decision to join the world of AVVID, be ready to go into it with patience and good planning. In the end, your employees will most likely travel less, be able to respond to meetings quicker, and actually be on time for them.

Solutions Fast Track

IP Telephony Multisite Centralized Call Processing Solutions

- ☑ This model provides consolidated VoIP management, which simplifies moves, adds, and changes.
- ☑ Because only one set of major devices is used, this reduces capital costs and the associated overhead of maintaining multiple devices.
- ☑ More disaster recovery and closer server management is required because now you have "all your eggs in one basket."
- ☑ Gives you better control of network resources since administrators can typically walk over to them for whatever maintenance is required.

IP Telephony Multisite Distributed Call Processing Solutions

- ☑ This model reduces WAN bandwidth requirements by keeping more of the processing local to each site.
- ☑ Also, this model can more easily withstand head office network issues such as virus attacks or errant router protocol problems.
- ☑ Even with the two preceding benefits, this model adds capital overhead, management, and additional WAN costs for each branch office which must now have a local network administrator.
- ☑ Sites can run more independently than a central solution, and thus act quicker to changing requirements of their own environment without waiting for the head office to react to their needs.

Multisite AVVID Solutions

- ☑ It can have dramatic cost savings over traditional training budgets.
- ☑ This model speeds information to the users by creating multiple avenues of data presentation.

- ☑ It also allows for remote mentoring of personnel without having associated travel costs.
- ☑ AVVID applications can provide interactive and automated customer support solutions, such as chat and whiteboarding solutions.

Frequently Asked Questions

The following Frequently Asked Questions, answered by the authors of this book, are designed to both measure your understanding of the concepts presented in this chapter and to assist you with real-life implementation of these concepts. To have your questions about this chapter answered by the author, browse to **www.syngress.com/solutions** and click on the **"Ask the Author"** form.

Q: What is centralized Voice over IP?

A: All servers and controlling functions are kept under tight control by the main office networks.

Q: How does distributed Voice over IP differ from the centralized solution?

A: Distributed CallManagers place the primary CallManager at the site of the managers or those making the changes and updates to the system, whereas a secondary CallManager is placed at each branch office.

Q: Why does the distributed solution require the most WAN bandwidth?

A: Site CallManagers must synchronize their CallManager SQL Server databases periodically, which also takes care of database cleanup and related operations.

Q: Why is a gatekeeper required for enterprise Voice over IP?

A: The gatekeeper controls call admission and acceptance for site-to-site communications, and guarantees that the available bandwidth is there for any of the WAN links, or it denies the call. This helps prevent congestion on the WAN links, and reduces the chance of poor quality calls.

Q: Can a primary CallManager synchronize its database across a dial-up ISDN link?

A: It depends on the speed of the link, but 128 Kbps is really the lower limit of what you'd want the WAN link speed to be for five to ten people.

Q: Can IP/TV replace HBO?

A: Yes and no. IP/TV is made for broadcast solutions where the destination is strictly receiving the data from any one of many sources, such as a TV studio or news camera. Outside services can be subscribed to, and then broadcast to others via the Cisco Broadcast Servers and Media Termination Points.

Q: How does IP Videoconferencing differ from Web cameras, or personal conferencing devices?

A: They don't, in reality. Both devices transmit voice and video, as well as receive voice and video from the other end. Each end of the link is an H.323 media point that compresses the information for transmission across just about any type of connection with a greater speed than 64 Kbps. Multiple group service might require as high as 768 Kbps link speeds to sustain dozens of sites using full motion video.

Appendix

Cisco AVVID and IP Telephony Design & Implementation Fast Track

This Appendix will provide you with a quick, yet comprehensive, review of the most important concepts covered in this book.

❖ Chapter 1: Old World Technologies

Introduction to PBXs

- ☑ Private Branch Exchange (PBX) systems provide corporate users with advanced voice services.
- ☑ The modern PBX is a reliable and robust tool on the network.
- ☑ Voice over IP (VoIP) technology is based on the Internet Protocol (IP) because it is the most common protocol in the networking world; however, the choice of this protocol brought with it problems of latency, queuing, and routing.
- ☑ Many PBXs today emulate the legacy key system's multiline presence, and this service is available with the current offering of AVVID.

Looking Inside the PBX

- ☑ The PBX uses trunks and lines to connect to resources.
- ☑ Call switching is distinct from call processing.
- ☑ Each PBX uses a variety of proprietary and standards-based protocols.

Interpreting PBX Terminology

- ☑ Bandwidths are based on analog channels: DS-0 (64 Kbps), DS-1 (T-1, 1.544 Mbps), DS-3 (45 Mbps).
- ☑ Links are called trunks.
- ☑ Some acronyms in the voice world have different meanings in the data world.

Working with Analog Systems

- ☑ Analog signals are continuous waveforms.
- ☑ Analog signals are susceptible to interference and are difficult to correct for errors.
- ☑ Analog signals cannot be compressed without loss.

Chapter 1 Continued

Benefiting from Digital Systems

- ☑ Digital signals are binary, made up of on or off signals.
- ☑ Digital signals can be compressed, corrected, and manipulated more easily than analog signals.
- ☑ Amplification can occur in digital signals without amplifying background noise and static.

Providing Video Services

- ☑ Video services can demand the most real-time bandwidth in the network.
- ☑ Video data is typically compressed to reduce its load on the network.
- ☑ One-to-many video is a perfect application for IP multicast.

❖ Chapter 2: New World Technologies

Introduction to IP Telephony

- ☑ Simplified administration is achieved by converging three separate networks into one, allowing one resource pool to administer the entire network.
- ☑ Toll bypass allows organizations to avoid costly telecommunications expenses by utilizing the data infrastructure.
- ☑ Unified messaging combines voice-mail, e-mail, and faxes into one easy-to-use interface.

IP Telephony Components

- ☑ CallManager provides the IP telephony network with a software-based PBX system.
- ☑ IP telephones provide the user interface to the IP telephony network.
- ☑ Gateways provide the interface between the IP telephony network and the public switched telephone network (PSTN) or a legacy PBX device.

Chapter 2 Continued

Exploring IP Telephony Applications

- ☑ WebAttendant replaces the traditional PBX attendant console.
- ☑ IP SoftPhone provides a software-based IP telephone handset.
- ☑ Third-party applications include software from Interactive Intelligence, Latitude, and ISI.

Introduction to Video

- ☑ Traditional video-conferencing utilizes ISDN lines in a point-to-point infrastructure.
- ☑ IP-based video-conferencing utilizes the H.323 specification allowing for video-conferencing over a variety of mediums.
- ☑ IP-based video-conferencing is much more efficient than traditional video-conferencing because the existing data infrastructure is utilized opposed to a separate infrastructure.
- ☑ Gateways provide access to the outside world from your internal network.
- ☑ Gatekeepers are used to permit or deny requests for video conferences.
- ☑ Multi-point control units (MCU) serve as a center for video-conferencing communications and infrastructure.

Enhancing Network Infrastructure

- ☑ Routers provide gateway services and voice aggregation for IP telephony by use of analog ports, FXO, FXS, E&M as well as digital trunking cards.
- ☑ Routers that support IP telephony include the 1751, 2600 Series, 3600 Series, and 7200 Series.
- ☑ Switches that support inline power modules include the 3524XL-PWR, 6000 Series, and 4000 Series.
- ☑ Inline power is also provided by using the Catalyst inline power patch panel.

What Does the Future Hold?

- ☑ Future revisions on CallManager include a call center solution.

Chapter 2 Continued

- ☑ Pizza box and integrated access devices will provide all-in-one functionality for branch offices.
- ☑ IOS-based versions of CallManager will further develop.

❖ Chapter 3: AVVID Gateway Selection

Introduction to AVVID Gateways

- ☑ In the Cisco AVVID world, there are voice and video gateways to provide connectivity to legacy networks. Cisco has voice gateways, which are standalone routers, IOS-based routers, and Catalyst switch-based routers.
- ☑ The standalone gateways include the DT-24+, DE-30+, and VG200. Router IOS-based gateway solutions are the 175x, 2600, 3600, 3810, 5300, 7200, and 7500. The switch-based gateways are the Catalyst 4000, 4200, and 6000 Series. These gateways run the following protocols: H.323, MGCP, Skinny, and SIP.
- ☑ The IP/VC 3500 family is the videoconferencing gateway products from Cisco.

Understanding the Capabilities of Gateway Protocols

- ☑ H.323 is the most supported gateway protocol, backed by the Cisco 1750, 2600, 3600, AS5300, 7200, and 7500 Series routers.
- ☑ Skinny Station Protocol allows a Skinny client to use TCP/IP to transmit and receive calls as with DT-24+, DE-30+, and VG200.
- ☑ MGCP is a master/slave protocol, where the gateway is the slave servicing commands from the master, which is the call agent. The MGCP protocol functions in an environment where the call control intelligence have been removed from the gateway.
- ☑ Session Initiation Protocol (SIP) is an application layer control protocol that can establish, modify, and terminate multimedia sessions or calls.

Chapter 3 Continued

Choosing a Voice Gateway Solution

- ☑ Determining the right voice gateway solutions will depend on a number of factors, from the size and scale of the organization to the budget.
- ☑ Solutions from a switch point-of-view would include, the Catalyst 4000, 4224/4248, and 6000 family. If you wish to use routers, you should choose from the following: the 1750, 2600, 3600, 3810, 7200, and 7500 Series. Access servers may be best in some instances, including the AS5300, the AS5400, and the AS5800. Cisco DT-24, DE-30, and VG-200 would suffice for standalone protocol solutions.
- ☑ For small- to mid-sized companies looking for a nice all-in-one solution, the ICS 7750, deployed with a Catalyst 3524XL-PWR switch and Cisco IP phones, would do wonderfully.
- ☑ The DPA 7610/7630 Voice Mail Gateway would be another important element of an AVVID solution. It provides a gateway allowing legacy voice mail systems to communicate with Cisco CallManagers.

Choosing a Video Gateway Solution

- ☑ Cisco's family of video gateway solutions can satisfy everyone from the small 40-person organization to those with 4000 employees.
- ☑ The IP/VC 3510 MCU connects three or more H.323 videoconference endpoints into a single multiparticipant meeting and is able to support ad-hoc or scheduled videoconferences. Participants can join by having the MCU dial to them or by using the Web interface.
- ☑ IP/VC 3520 and 3525 gateways provide the translation services between H.320 and H.323 networks. This system allows users to conduct videoconferencing across the IP LAN, or via the PSTN. The IP/VC 352x series gateways support V.35, ISDN BRI, and ISDN PRI interfaces. IP/VC 3530 VTA translates from a H.320 ISDN-based system to a H.323 IP-based network. The IP/VC 3540 solution is a highly scalable MCU, which is chassis-based and expands to up to three modules. These modules come in 30-, 60-, and 100-user versions.

Chapter 3 Continued

Multimedia Conference Manager Services

- ☑ Multimedia Conference Manager (MCM) works in conjunction with Cisco's IP/VC products, and services a H.323 gatekeeper and proxy.
- ☑ MCM is a part of the Cisco IOS for the following router platforms: 2500, 2600, 3600, 3810, and 7200.
- ☑ The MCM gatekeeper functions include: zone administration, RAS, AAA services, bandwidth management, session management, and call accounting. The proxy service provides QoS capabilities to the videoconferencing sessions.

❖ Chapter 4: AVVID Clustering

CallManager Clustering

- ☑ Cisco AVVID infrastructure includes a variety of features to facilitate load balancing, scalability, and redundancy for IP telephony and multimedia conference solutions.
- ☑ Cisco CallManager clusters are used to improve the scalability and reliability of Cisco IP telephony solutions.
- ☑ Multipoint Control Unit cascading is used to improve the scalability of voice/video conferencing.
- ☑ A maximum of eight Cisco CallManagers can be members of a cluster, with as many as six used for call processing.
- ☑ The possible roles of servers within a cluster are: database publisher server, TFTP server, application server, primary call-processing server, and backup call-processing server.
- ☑ Intra-cluster communications rely on high-speed network connections, and are not supported across WANs.
- ☑ The CallManager database contains the configuration of all IP telephony devices.
- ☑ Real-time data replicated between servers in a cluster consists of registration information of IP telephony devices.
- ☑ Many CallManager features do not function between different clusters.

Chapter 4 Continued

- ☑ Database redundancy is achieved by replicating the publisher database to all servers within a cluster.
- ☑ Redundancy groups facilitate server failover. A device is associated with a redundancy group, which is a list of up to three servers. If the primary server fails, call processing is transferred to the secondary server.
- ☑ Balanced call processing can be achieved by assigning different primary servers to different groups of devices.
- ☑ Device weights are used to calculate the maximum number of devices that can be supported by a single CallManager server.

Video Clustering

- ☑ A maximum of 15 conference participants can be supported by a single MCU.
- ☑ Two or more MCUs can be cascaded to support larger conferences.
- ☑ Conference participants are unaware of the cascaded nature of the conference.
- ☑ Only a single voice/video data stream exists between cascaded MCUs.
- ☑ Voice/video traffic can be localized by correctly dispersing MCUs across a network.
- ☑ The number of MCUs that can be cascaded together depends on available bandwidth.
- ☑ To invite a MCU to join a conference from a terminal, dial the host conference password, the invite code **, followed by the conference password of the invited MCU.

❖ Chapter 5: Voice and Video Gatekeeper Design

Understanding Gatekeeper Basics

- ☑ A gatekeeper is a central point of control for an H.323 (voice and video) network.

Chapter 5 Continued

- ☑ Gatekeepers usually use E.164 addressing (telephone numbers) for identifying endpoints and routing calls within a network.
- ☑ Gatekeepers run an H.323/MCM feature set IOS on many common Cisco routers.

A Gatekeeper's Role in Voice and Video Networking

- ☑ Gatekeepers manage one or multiple zones and permit or reject calls into or out of each zone.
- ☑ Gatekeepers can provide accounting information for calls, such as length of call, time of call, number called, and so on.
- ☑ Cisco's Multimedia Conference Manager (MCM) can act as a proxy for increased security and QoS as well as a gatekeeper.
- ☑ Video gatekeepers can be embedded in the video controller or can be an MCM.
- ☑ Video gatekeepers interface with gateways for off-network calls, such as ISDN videoconferences.
- ☑ Gatekeepers monitor (and limit) bandwidth usage to assure existing calls receive high quality.

❖ Chapter 6: DSPs Explained

DSP Provisioning

- ☑ The Cisco DSP module is a Texas Instruments model C542 and C549 72-pin SIMM. These DSPs work with two levels of CODEC complexity: medium and high.
- ☑ The medium-complexity CODECs that work with the Cisco DSP are G.711 (a-law and μ-law), G.726, G.729a, G.729ab, and Fax-relay. The high-complexity CODECs include the G.728, G.723, G.729, G.729b, and Fax-relay.
- ☑ The DSP resources are used for conference bridging and transcoding.

Chapter 6 Continued

Conferencing and Transcoding

- ☑ Conferencing is the process of joining multiple callers into a single multiway call. The two types of multiparticipant voice calls supported by the Cisco CallManager are ad-hoc and meet-me.
- ☑ DSP resources are used in the conference bridge scenario to convert VoIP calls into TDM streams and sum them into a single call.
- ☑ Transcoding is the process of converting IP packets of voice streams between a low bit-rate (LBR) CODEC to G.711. Transcoding functions can be done by converting G.723 and G.729 CODECs to G.711.
- ☑ Conferencing and transcoding is performed either by hardware or software. The software version is performed on a Cisco CallManager server, while the hardware solutions are the Catalyst 4000 AGM module, Catalyst 6000 8-port T1/E1Voice and Services module, and NM-HDV module.

Catalyst 4000 Modules

- ☑ The Catalyst 4000 Access Gateway Module (AGM) provides voice network services to the Catalyst 4000 switch, VoIP IP WAN routing, and an IP telephony mode for use with a voice gateway. The Catalyst 4000 AGM supports voice interface cards (VICs) and WAN interface cards (WICs) from the 1600/1700/2600/3600 Series routers.

Catalyst 6000 Modules

- ☑ The Catalyst 6000 switch module features an 8-port Voice T1/E1 and Services module, WS-X6608-E1 or WS-X6608-T1.
- ☑ The Voice T1/E1 module supports T1/E1 CCS signaling, ISDN PRI network, and user-side signaling. Similar to the AGM module for the Catalyst 4000, the Voice T1/E1 can be provisioned for conferencing and transcoding. The Voice T1/E1 can do mixed CODEC conferencing, whereas the AGM only does G.711 conferencing with individual DSP resources.

NM-HDV Modules

- ☑ The biggest benefit of this module is PBX leased line replacement and toll bypass, meaning that a company's long distance expenses can all but be

Chapter 6 Continued

eliminated. Platform support includes VG200, 2600, 3600, and Catalyst AGM E1 Models (medium complexity involving NM-HDV-1E1-12, NM-HDV-1E1-30, and NM-HDV-2E1-60). With E1 Models (high complexity M-HDV-1E1-30E), or T1 Models, and medium complexity (NM-HDV-1T1-12, NM-HDV-1T1-24, and NM-HDV-2T1-48) supported, it will also support T1 Models (high complexity NM-HDV-1T1-24E).

Sample Design Scenarios

- ☑ When designing your DSP provisioning, you must take into account the number of users, the type of applications using different CODEC, and the overall IP telephony design to determine which solution best fits your needs, whether it's using the CallManager itself or one of the Catalyst switches.
- ☑ The branch office environment is an excellent candidate for the Catalyst 4000 switch with an Access Gateway module (AGM). This solution can provide 10/100/1000 Ethernet switching with inline power for IP phones, PSTN connectivity, IP routing, and also serve as a DSP resource. The DSP resources provide conferencing and transcoding services for your user population.
- ☑ The enterprise campus has higher scalability requirements than the branch office. With this in mind, you should consider the Catalyst 6000 with the 8-port T1/E1 Voice and Service module as a good fit for the needs of this environment.

❖ Chapter 7: AVVID Applications

Creating Customer Contact Solutions

- ☑ Make sure you understand the customer's needs.
- ☑ Provide the client with the solution that best suits these needs.
- ☑ Make sure to stay within the Cisco recommended guidelines.
- ☑ With the IP contact center, there are many different components. Make sure the version numbers needed to run the solution are all compatible.

Chapter 7 Continued

Providing Voice Recording Options

- ☑ Make sure the infrastructure can support voice recording.
- ☑ Define the endpoints that need to be recorded, and implement a policy using this as a framework.

Call Accounting, Billing, and Network Management Solutions

- ☑ Understand the requirements in enabling CDRs throughout your network, not just on the Cisco CallManager, but also on your router infrastructure (if possible).
- ☑ Look at the Administrative Reporting Tool (ART) with Cisco CallManager to decide whether this would provide you with the information needed before looking at external solutions.
- ☑ Define the information needed with your reports, and based on this, look for solutions that meet the requirement you and your customers have.

Designing Voice and Unified Messaging Solutions

- ☑ Decide on the version of Unity needed.
- ☑ If upgrading from voice mail to unified messaging, do not forget the possible hardware requirements.
- ☑ You should be running Microsoft Exchange 5.5 or Exchange 2000, with future support for other platforms.

Understanding Other Voice Applications

- ☑ Keep it as simple as possible, if services or applications are not needed, do not enable them. It complicates the configuration.
- ☑ IP Automated Attendant (AA) is extremely useful in large organizations where switchboard operators are normally overworked. Automated Attendant, as its name suggests, provides automated functions an attendant might normally perform.

Chapter 7 Continued

☑ WebAttendant is a Web-based graphical user interface (GUI) that works with a standard Web browser without making any changes to the browser itself. The only thing needed for the installation is to download the application from the Cisco CallManager Install Plug-ins page.

❖ Chapter 8: Advanced QoS for AVVID Environments

Using the Resource Reservation Protocol

☑ RSVP does not provide QoS directly to applications, but instead, coordinates an overall service level by making reservation requests across the network. It is up to other QoS mechanisms to actually prevent and control congestion, provide efficient use of links, and classify and police traffic.

☑ End-to-end resource reservation can only be accomplished by using RSVP on every router end-to-end, but it is not mandatory that RSVP be enabled everywhere on a network. RSVP has the built-in capability to tunnel over non-RSVP aware nodes.

☑ Because of the resources required for each reservation, RSVP has some distinct scaling issues that make it doubtful it will ever be implemented successfully on a very large network, or on the Internet, in its current revision.

Using Class-Based Weighted Fair Queuing

☑ CBWFQ carries the WFQ algorithm further by allowing user-defined classes, which allow greater control over traffic queuing and bandwidth allocation.

☑ Flow-based WFQ automatically detects flows based on characteristics of the third and fourth layers of the OSI model. Conversations are singled out into flows by source and destination IP address, port number, and IP precedence.

☑ CBWFQ allows the creation of up to 64 individual classes plus a default class. The number and size of the classes are, of course, based on the bandwidth. By default, the maximum bandwidth that can be allocated to user-defined classes is 75 percent of the link speed.

Chapter 8 Continued

Using Low Latency Queuing

- ☑ LLQ creates a strict priority queue that you can think of as resting on top of all other CBWFQ queues.
- ☑ LLQ overcomes the fact that low latency transmission may not be provided to packets in congestion situations, since all packets are transmitted fairly, based on their weight.
- ☑ Because of the nature of the LLQ, it is recommended that only voice traffic be placed in that queue.

Using Weighted Random Early Detection

- ☑ RED works on the basis of active queue management, and addresses the shortcomings of tail drop.
- ☑ WRED was primarily designed for use in IP networks dominated by TCP, because UDP traffic is not responsive to packet drop like TCP.
- ☑ WRED treats non-IP traffic as precedence 0, the lowest precedence. Therefore, non-IP traffic will be lumped into a single bucket and is more likely to be dropped than IP traffic. This may cause problems if most of your important traffic is something other than IP.

Using Generic Traffic Shaping and Frame Relay Traffic Shaping

- ☑ FRTS and GTS both use a token bucket, or credit manager, algorithm to service the main queuing mechanism and send packets out the interface. FRTS is commonly used to overcome data-rate mismatches.
- ☑ FRTS and GTS act to limit packet rates sent out an interface to a mean rate, while allowing for buffering of momentary bursts.
- ☑ Recall that queuing mechanisms will only kick in when there is congestion, so we need a mechanism to create congestion at the head-end. This is a common need on Frame Relay networks where the home office has much more bandwidth than any individual remote office.

Chapter 8 Continued

Running in Distributed Mode

- ☑ When a process is run on the VIP instead of the main processor, the service is said to be running in *distributed mode.*
- ☑ Most of the QoS features you will find useful in an AVVID environment were introduced (in distributed mode) in 12.1(5)T.

Using Link Fragmentation and Interleaving

- ☑ Real-time streams usually consist of small packets, and jitter is caused when the regularly timed transmission of these packets is interrupted by the serialization delay of sending a large packet. Serialization delay is the fundamental time it takes a packet to be sent out a serial interface.
- ☑ Using a feature like LLQ or PQ can significantly reduce delays on real-time traffic, but even with this enabled, the time a real-time packet may have to wait for even one large packet to be transmitted could be large enough to add jitter to the stream.
- ☑ Link Fragmentation and Interleaving overcomes this by reducing the maximum packet size of all packets over a serial link to a size small enough that no single packet will significantly delay critical real-time data.

Understanding RTP Header Compression

- ☑ RTP encapsulates UDP and IP headers, and the total amount of header information (RTP/UDP/IP) adds up to 40 bytes. Since small packets are characteristic of multimedia streams, that is a lot of overhead. Most of the header information does not change from packet to packet, so RTP header compression can reduce this 40-byte header to about 5 bytes on a link-by-link basis.
- ☑ RTP header compression can be useful on any narrowband link. Narrowband is usually defined by speeds less than T1.
- ☑ Since cRTP is performed by the main processor, enabling it could cause your CPU utilization to jump if you have high packet rates, lots of serial interfaces, or large serial interfaces. Use this feature with caution.

❖ Chapter 9: AVVID Dial Plans

What Is a Dial Plan?

- ☑ Configuring dial peers for use is essential when designing and implementing Voice over IP on your network. Dial peers identify the calling source and the destination points so as to define what attributes are assigned to each call.
- ☑ Configuring a dial peer for POTS can help you shape the deployment of your dial peers.
- ☑ By configuring VoIP dial peers, you can enable the router to make outbound calls to other telephony devices located within the network.
- ☑ Dial peers for inbound and outbound calls are used to receive and complete calls. You must remember that the definition of inbound and outbound is based on the perspective of the router. What this means is that a call coming into the router is considered an inbound call and a call originating from the router is considered an outbound call.
- ☑ To associate a dialed string with a specific telephony device, you would use the destination pattern. With it, the dialed string will compare itself to the pattern and then will be routed to the voice port or the session target (discussed later) voice network dial peer. If the call is an outbound call, the destination pattern could also be used to filter the digits that will be forwarded by the router to the telephony device or the PSTN. A destination pattern must be configured for each and every POTS and VoIP dial peer configured on the router.
- ☑ The session target is the IP address of the router to which the call will be directed once the dial peer is matched.
- ☑ Route patterns (on-net) allow you to connect to multiple sites across a WAN with connections like frame or dedicated circuits using available network resources.
- ☑ With Cisco CallManager, you are able to create route patterns that allow you to route calls that differentiate between local calls and long distance calls.

Chapter 9 Continued

Cisco CallManager Dial Plans

- ☑ By using Cisco CallManager, you can allow for greater growth and functionality within your network because it was designed to be integrated with IOS gateways.
- ☑ The creation of dial plans for internal calls to IP phones are registered within a Cisco CallManager cluster.
- ☑ External calls use a route pattern to direct off-network calls to a PSTN gateway. Route patterns can also be used if there are Cisco CallManagers located on a WAN-connected network.
- ☑ A route pattern is the addressing method that identifies the dialed number and uses route lists and route group configurations to determine the route for call completion.
- ☑ Digit manipulation involves digit removal and prefixes, digit forwarding, and number expansion.
- ☑ Route lists are configured to map the routes of a call to one or more route groups.
- ☑ Route groups allow you to control telephony devices.
- ☑ Telephony devices are any devices capable of being connected to a route group.
- ☑ the digit translation table manipulates dialed digits and is supported within Cisco Call Manager
- ☑ Fixed-length dial peers versus Variable-length dial peers—This will help you to decide what to use in your network.
- ☑ Two-stage dialing occurs when a voice call is destined for the network, and the router placing the call collects all of the dialed digits.

Creation of Calling Restrictions and Configuration of Dial Plan Groups

- ☑ Within Cisco CallManager, you can create calling restrictions per each telephony device, or create closed dial plan groups (as long as they fall within the same Cisco CallManager). What this means is that users residing within the same Cisco CallManager can be grouped together with the same

calling restrictions and dial plans. For example if you have development teams that need to talk to only each other, you can restrict their dial plans to within the group, or limit their ability to call long distance.

- ☑ A partition is a group of telephony devices that have similar reach ability. These devices are composed of route patterns, IP SoftPhones, directory numbers, and so on.
- ☑ A calling search space is a list of partitions that can be accessed by users in order to place a call. These calling search spaces are only allocated to telephony devices that can start calls. With these calling search spaces implemented, it is simple to create and use dialing restrictions, because users are only allowed to dial those partitions in the calling search space they are assigned to. If the user tries to call outside the allowed partitions, they will receive a busy signal.
- ☑ The combination of partitions and calling search spaces can allow autonomous dial ranges on a partition basis. Extension and access codes located within different partitions can have overlapping number schemes, and will still work independently of each other. This is usually seen in the implementation of a centralized call processing system. In this example, all sites that use the same Cisco CallManager can dial the number 9 to access the PSTN, even if they are located on different WAN segments.

Guidelines for the Design and Implementation of Dial Plans

- ☑ As with any project, its complexity will depend on the number of variables factored in. Dial plan complexity can vary, based on any number of configuration choices, such as the total amount of paths a call can be sent through.
- ☑ When configuring single-site campuses, you will often implement a simple dial plan that can provide intraoffice calling (with four or five digits depending on the site) and connections to the PSTN (usually by dialing a 9). Long distance would also be handled by the PSTN with the dialing party using a 9, then a 1, and the area code before dialing the seven-digit number.
- ☑ When you go to implement AVVID, you should work under the assumption that the less complex it is, the better. Find out what is used on a normal

Chapter 9 Continued

(daily) basis, and what features are seldom used. With these answers, you can create a plan that meets the needs of the client.

- ☑ Based on the assumption that this will be a Cisco IOS-based H.323 gateway, you would then point the router POTS dial peer to the PSTN port (or ports) and use a destination pattern of "9" to match the leading digit that will come from the Cisco CallManager. The match on the "9" will make the dial peer remove the 9, so the rest of the number is passed.
- ☑ When creating a dial plan for a multisite WAN, you must have sufficient resources to make it function properly. If you don't have the proper link bandwidth, the call will always route over the PSTN, negating the benefits that multisite WAN connections are supposed to give you.

The Role and Configuration of a Cisco CallManager and Gatekeeper

- ☑ By implementing H.323 gatekeepers for admission control, you can control the number of calls allowed to and from specific areas. This will assist you in the management of bandwidth and resources for your sites and overall infrastructure. The Cisco CallManager uses the gatekeeper to perform admission control, especially in infrastructures that use hub and spoke architecture for network centralization.
- ☑ The Cisco Call Manager dial plan model requires that all Cisco CallManagers located within a cluster be connected through an intercluster trunk with a route pattern for each of the other clusters within the domain.
- ☑ The Gatekeeper dial plan model helps to clean up the overhead inherent in the Cisco CallManager model. This is because the Cisco CallManager only needs to maintain one intercluster trunk, known as the "anonymous device." This "device" is like a point-to-multipoint connection in frame relay, as the Cisco CallManagers don't need to be fully meshed. In this setup, the gatekeeper is able to use the anonymous device to route calls through the network to the correct Cisco CallManager (or cluster).
- ☑ The Hybrid model allows for the automatic overflow to the PSTN of calls destined for the WAN which are unable to allocate sufficient resources. It only needs one anonymous device for each Cisco CallManager (cluster),

thus minimizing the overhead of having to mesh the Cisco CallManagers. It does require two routes for each destination, however, one to the WAN and one to the PSTN. The drawback is you need to configure the dial plan on the gatekeeper and the Cisco CallManager.

- ☑ For every gatekeeper located within your domain, you must configure the intercluster CODEC you would like to use, as well as enable the anonymous device. When that is complete, you will need to configure the router pattern to allow calls between clusters. You would do this by selecting a CODEC for all intercluster calls, defining the region that the gatekeeper and cluster are located in, and select the appropriate compression rate.
- ☑ When configuring the Cisco CallManager gatekeeper, you are required to enter a zone. Each Cisco CallManager will register with that zone, its zone prefix (the directory number ranges), the bandwidth allowed for each call admission, and the technology prefix for voice-enabled devices. Cisco CallManager will need the gatekeeper to explicitly specify the IP address of the Cisco CallManager within a single zone, then you must disable the registration of all other IP address ranges so it can only exist within that zone.

Video Dial Plan Architecture

- ☑ Corporate video conferencing was first introduced in the 1980's as a way to help people in different cities communicate more effectively. These first-generation solutions were based on the ITU H.320 standards defining ISDN connection-based videoconferencing.
- ☑ The Cisco Multimedia Conference Manager (Cisco MCM) is a specialized Cisco IOS software image that lets network administrators support H.323 applications on their networks without compromising mission-critical traffic from other applications. The Cisco MCM serves two main functions: it acts as a gatekeeper, and as a proxy.
- ☑ A gateway is an optional element that can be implemented within the H.323 deployment. It is an endpoint on the LAN that can provide real-time, two-way communication between H.323 terminals or other gateways. It is also capable of using the LAN and other ITU terminals located on the WAN by using H.425 and Q.931 protocols.
- ☑ A proxy gateway is a secured connection between H.323 sessions. The Cisco Multimedia Conference Manager contains a proxy as part of its

Chapter 9 Continued

infrastructure so it can provide QoS, traffic shaping, and security and policy management for H.323 traffic across any secured connection.

- ☑ The H.323 gatekeeper is an optional component capable of providing call control services to H.323 endpoints. You may implement multiple gatekeepers within your network, and they will remain logically separate from the endpoints. There are currently no standards for gatekeeper-to-gatekeeper communications, so you may want to explore other options before installing multiple gatekeepers within the same segment. You could install terminals, MCUs, gateways, or other non-H.323 LAN devices since these may coexist in the same environment.
- ☑ An MCU is a device that aids in getting calls to three or more endpoints in conference type deployments. It is usually a centralized device that assists in the facilitation of conference sessions for data, video, and/or audio.
- ☑ Video dial peers is a feature supported only on the MC3810 Multi-Service Concentrator.

❖ Chapter 10: Designing and Implementing Single Site Solutions

Using AVVID Applications in IP Telephony Single Site Solutions

- ☑ Single site VoIP systems can be a cost-effective replacement for traditional PBX systems, especially in locations where available PBX solutions are limited. This is most helpful in places where you have more network engineers capable of managing Cisco devices than traditional telephony solutions.
- ☑ VoIP permits easy remote management of the entire system via CallManager's Web interface. Even the server's services can be stopped and restarted by way of the Web interface.
- ☑ By using the inline power enterprise model of switches, the customer can future-proof growth needs for both voice and data applications, foregoing the need for replacement devices and the consequent disruption of existing services.

Chapter 10 Continued

Using AVVID Applications in Single Site Solutions

- ☑ With the development of the Unity product, Cisco provides great messaging capability that finally breaks all ties to traditional telephony systems. Now, full deployment of AVVID solutions can be achieved to other sites by using only external WAN communications, as well as all internal communications riding on the Cisco-powered enterprise.
- ☑ Because Unity integrates with Exchange Server, and uses the native Exchange directory services, it is easy to deploy and manage, and has the flexibility to handle various messaging needs. Unity works with all standards-based SMTP, POP3, and IMAP4 clients, maintaining ease of use and portability between clients.
- ☑ CallManager provides excellent flexibility for moves, adds, and changes. Its Web interface makes the system accessible from any location, even from dial-up modems with slow speeds. CallManager is highly extensible, allowing it to serve thousands of users in a centralized or distributed environment.

Using AVVID Applications in Video Single Site Solutions

- ☑ Cisco video solutions offer dramatic savings in the area of training by dramatically reducing or even eliminating travel costs. Presentations can be shipped to the site when so desired, and easily deployed.
- ☑ The flexibility to present video on demand speeds information to users whenever needed. Video on demand (VOD) means users can come back from vacation and review that missed presentation from the head office without needing to schedule a new briefing.
- ☑ Video solutions allow for remote mentoring at any time, by anyone. New personnel no longer have to fly to the head office for indoctrination, nor do they have to wait for the next session. Trainers can also create their own labs and exercises where the experts reside, without any travel costs. The new videos can then be shared at any location.

❖ Chapter 11: Designing and Implementing Multisite Solutions

IP Telephony Multisite Centralized Call Processing Solutions

- ☑ This model provides consolidated VoIP management, which simplifies moves, adds, and changes.
- ☑ Because only one set of major devices is used, this reduces capital costs and the associated overhead of maintaining multiple devices.
- ☑ More disaster recovery and closer server management is required because now you have "all your eggs in one basket."
- ☑ Gives you better control of network resources since administrators can typically walk over to them for whatever maintenance is required.

IP Telephony Multisite Distributed Call Processing Solutions

- ☑ This model reduces WAN bandwidth requirements by keeping more of the processing local to each site.
- ☑ Also, this model can more easily withstand head office network issues such as virus attacks or errant router protocol problems.
- ☑ Even with the two preceding benefits, this model adds capital overhead, management, and additional WAN costs for each branch office which must now have a local network administrator.
- ☑ Sites can run more independently than a central solution, and thus act quicker to changing requirements of their own environment without waiting for the head office to react to their needs.

Multisite AVVID Solutions

- ☑ It can have dramatic cost savings over traditional training budgets.
- ☑ This model speeds information to the users by creating multiple avenues of data presentation.

Chapter 11 Continued

- ☑ It also allows for remote mentoring of personnel without having associated travel costs.
- ☑ AVVID applications can provide interactive and automated customer support solutions, such as chat and whiteboarding solutions.

Index

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Global Knowledge™

Train with Global Knowledge

The right content, the right method, delivered anywhere in the world, to any number of people from one to a thousand. Blended Learning Solutions™ from Global Knowledge.

Train in these areas:

Network Fundamentals
Internetworking
A+ PC Technician
WAN Networking and Telephony
Management Skills
Web Development
XML and Java Programming
Network Security
UNIX, Linux, Solaris, Perl
Cisco
Enterasys
Entrust
Legato
Lotus
Microsoft
Nortel
Oracle

www.globalknowledge.com

Global Knowledge™

Every hour, every business day
all across the globe
*Someone just **like you***
is being trained by
Global Knowledge.

Only Global Knowledge offers so much content in so many formats—Classroom, Virtual Classroom, and e-Learning. This flexibility means Global Knowledge has the IT learning solution you need.

Being the leader in classroom IT training has paved the way for our leadership in technology-based education. From CD-ROMs to learning over the Web to e-Learning live over the Internet, we have transformed our traditional classroom-based content into new and exciting forms of education.

Most training companies deliver only one kind of learning experience, as if one method fits everyone. Global Knowledge delivers education that is an exact reflection of you. No other technology education provider integrates as many different kinds of content and delivery.

www.globalknowledge.com

Blended Learning Solutions™ from Global Knowledge

The Power of Choice is Yours.

Get the IT Training you need— how and when you need it.

Mix and match our Classroom, Virtual Classroom, and e-Learning to create the exact blend of the IT training you need. You get the same great content in every method we offer.

Self-Paced e-Learning

Self-paced training via CD or over the Web, plus mentoring and Virtual Labs.

Virtual Classroom Learning

Live training with real instructors delivered over the Web.

Classroom Learning

Train in the classroom with our expert instructors.

1-800-COURSES **www.globalknowledge.com**

Global Knowledge™

9000 Regency Parkway, Suite 500
Cary, NC 27512
1-800-COURSES
www.globalknowledge.com

At Global Knowledge, we strive to support the multiplicity of learning styles required by our students to achieve success as technical professionals. We do this because we know our students need different training approaches to achieve success as technical professionals. That's why Global Knowledge has worked with Syngress Publishing in reviewing and recommending this book as a valuable tool for successful mastery of this subject.

As the world's largest independent corporate IT training company, Global Knowledge is uniquely positioned to recommend these books. The first hand expertise we have gained over the past several years from providing instructor-led training to well over a million students worldwide has been captured in book form to enhance your learning experience. We hope the quality of these books demonstrates our commitment to your lifelong learning success. Whether you choose to learn through the written word, e-Learning, or instructor-led training, Global Knowledge is committed to providing you the choice of when, where and how you want your IT knowledge and skills to be delivered. For those of you who know Global Knowledge, or those of you who have just found us for the first time, our goal is to be your lifelong partner and help you achieve your professional goals.

Thank you for the opportunity to serve you. We look forward to serving your needs again in the future.

Warmest regards,

Duncan M. Anderson
President and Chief Executive Officer, Global Knowledge

P.S. Please visit us at our Web site www.globalknowledge.com.

Enter the Global Knowledge Chrysler PT Cruiser Sweepstakes

This sweepstakes is open only to legal residents of the United States who are Business to Business MIS/IT managers or staff and training decision makers, that are 18 years of age or older at time of entry. Void in Florida & Puerto Rico.

OFFICIAL RULES

No Purchase or Transaction Necessary To Enter or Win, purchasing will not increase your chances of winning.

1. How to Enter: Sweepstakes begins at 12:00:01 AM ET May 1, 2001 and ends 12:59:59 PM ET December 31, 2001 the ("Promotional Period"). There are four ways to enter to win the Global Knowledge PT Cruiser Sweepstakes: Online, at Trade shows, by mail or by purchasing a course or software. Entrants may enter via any of or all methods of entry.

[1] To be automatically entered online, visit our web at www.globalknowledge.com click on the link named Cruiser and complete the registration form in its entirety. All online entries must be received by 12:59:59 PM ET December 31, 2001. Only one online entry per person, per e-mail address. Entrants must be the registered subscriber of the e-mail account by which the entry is made.

[2] At the various trade shows, during the promotional period by scanning your admission badge at our Global Knowledge Booth. All entries must be made no later than the close of the trade shows. Only one admission badge entry per person.

[3] By mail or official entry blank available at participating book stores throughout the promotional period. Complete the official entry blank or hand print your complete name and address and day & evening telephone # on a 3"x5" card, and mail to: Global Knowledge PT Cruiser Sweepstakes, P.O. Box 4012 Grand Rapids, MN 55730-4012. Entries must be postmarked by 12/31/01 and received by 1/07/02. Mechanically reproduced entries will not be accepted. Only one mail in entry per person.

[4] By purchasing a training course or software during the promotional period: online at http://www.globalknowledge.com or by calling 1-800-COURSES, entrants will automatically receive an entry onto the sweepstakes. Only one purchase entry per person.

All entries become the property of the Sponsor and will not be returned. Sponsor is not responsible for stolen, lost, late, misdirected, damaged, incomplete, illegible entries or postage due mail.

2. Drawings: There will be five [5] bonus drawings and one [1] prize will be awarded in each bonus drawing. To be eligible for the bonus drawings, on-line entries, trade show entries and purchase entries must be received as of the dates listed on the entry chart below in order to be eligible for the corresponding bonus drawing. Mail in entries must be postmarked by the last day of the bonus period, except for the month ending 9/30/01 where mail in entries must be postmarked by 10/1/01 and received one day prior to the drawing date indicated on the entry

chart below. Only one bonus prize per person or household for the entire promotion period. Entries eligible for one bonus drawing will not be included in subsequent bonus drawings.

Bonus Drawings	Month starting/ending 12:00:01 AM ET/11:59:59 PM ET	Drawing Date on or about
1	5/1/01-7/31/01	8/8/01
2	8/1/01-8/31/01	9/11/01
3	9/1/01-9/30/01	10/10/01
4	10/1/01-10/31/01	11/9/01
5	11/1/01-11/30/01	12/11/01

There will also be a grand prize drawing in this sweepstakes. The grand prize drawing will be conducted on January 8, 2002 from all entries received. Bonus winners are eligible to win the Grand prize.

All random sweepstakes drawings will be conducted by Marden-Kane, Inc. an independent judging organization whose decisions are final. All prizes will be awarded. The estimated odds of winning each bonus drawing are 1:60,000, for the first drawing and 1:20,000 for the second, third, fourth and fifth drawings, and the estimated odds of winning the grand prize drawing is 1:100,000. However the actual odds of winning will depend upon the total number of eligible entries received for each bonus drawing and grand prize drawings.

3. Prizes: Grand Prize: One (1) PT Cruiser 2002 model Approx. Retail Value (ARV) $18,000. Winner may elect to receive the cash equivalent in lieu of the car. Bonus Prizes: Five (5), awarded one (1) per bonus period. Up to $1,400.00 in self paced learning products ARV up to $1,400.00 each.

No substitutions, cash equivalents, except as noted, or transfers of the prize will be permitted except at the sole discretion of the Sponsor, who reserves the right to substitute a prize of equal or greater value in the event an offered prize is unavailable for any reason. Winner is responsible for payment of all taxes on the prize, license, registration, title fees, insurance, and for any other expense not specifically described herein. Winner must have and will be required to furnish proof of a valid driver's license. Manufacturers warranties and guarantees apply.

4. Eligibility: This sweepstakes is open only to legal residents of the United States, except Florida and Puerto Rico residents who are Business to Business MIS/IT managers or staff and training decision makers, that are 18 years of age or older at the time of entry. Employees of Global Knowledge Network, Inc and its subsidiaries, advertising and promotion agencies including Marden-Kane, Inc., and immediate families (spouse, parents, children, siblings and their respective spouses) living in the same household as employees of these organizations are ineligible. Sweepstakes is void in Florida and Puerto Rico and is subject to all applicable federal, state and local laws and regulations. By participating, entrants agree to be bound by the official rules and accept decisions of judges as final in all matters relating to this sweepstakes.

5. Notification: Winners will be notified by certified mail, return receipt requested, and may be required to complete and sign an Affidavit of Eligibility/Liability Release and, where legal, a Publicity Release, which must be returned, properly executed, within fourteen (14) days of

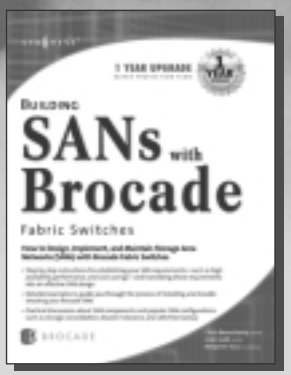
issuance of prize notification. If these documents are not returned properly executed or are returned from the post office as undeliverable, the prize will be forfeited and awarded to an alternate winner. Entrants agree to the use of their name, voice and photograph/likeness for advertising and promotional purposes for this and similar promotions without additional compensation, except where prohibited by law.

6. Limitation of Liability: By participating in the Sweepstakes, entrants agree to indemnify and hold harmless the Sponsor, Marden-Kane, Inc. their affiliates, subsidiaries and their respective agents, representatives, officers, directors, shareholders and employees (collectively, "Releasees") from any injuries, losses, damages, claims and actions of any kind resulting from or arising from participation in the Sweepstakes or acceptance, possession, use, misuse or nonuse of any prize that may be awarded. Releasees are not responsible for printing or typographical errors in any instant win game related materials; for stolen, lost, late, misdirected, damaged, incomplete, illegible entries; or for transactions, or admissions badge scans that are lost, misdirected, fail to enter into the processing system, or are processed, reported, or transmitted late or incorrectly or are lost for any reason including computer, telephone, paper transfer, human, error; or for electronic, computer, scanning equipment or telephonic malfunction or error, including inability to access the Site. If in the Sponsor's opinion, there is any suspected or actual evidence of electronic or non-electronic tampering with any portion of the game, or if computer virus, bugs, unauthorized intervention, fraud, actions of entrants or technical difficulties or failures compromise or corrupt or affect the administration, integrity, security, fairness, or proper conduct of the sweepstakes the judges reserve the right at their sole discretion to disqualify any individual who tampers with the entry process and void any entries submitted fraudulently, to modify or suspend the Sweepstakes, or to terminate the Sweepstakes and conduct a random drawing to award the prizes using all non-suspect entries received as of the termination date. Should the game be terminated or modified prior to the stated expiration date, notice will be posted on http://www.globalknowledge.com. Any attempt by an entrant or any other individual to deliberately damage any web site or undermine the legitimate operation of the promotion is a violation of criminal and civil laws and should such an attempt be made, the sponsor reserves the right to seek damages and other remedies from any such person to the fullest extent permitted by law. Any attempts by an individual to access the web site via a bot script or other brute force attack or any other unauthorized means will result in the IP address becoming ineligible. Use of automated entry devices or programs is prohibited.

7. Winners List: For the name of the winner visit our web site www.globalknowledge.com on January 31, 2002.

8. Sponsor: Global Knowledge Network, Inc., 9000 Regency Parkway, Cary, NC 27512. Administrator: Marden-Kane, Inc. 36 Maple Place, Manhasset, NY 11030.

SYNGRESS SOLUTIONS...



AVAILABLE NOW
ORDER at
www.syngress.com

Building SANs with Brocade Fabric Switches

As a superior alternative to traditional direct-attached storage models, Storage Area Networks (SANs) have fundamentally changed the way enterprise storage infrastructure is designed and built. Moreover, SANs have enabled a wide range of new applications and competitive advantages. To explain these benefits, Brocade Communications Systems and Syngress Publishing bring you Building SANs with Brocade Fabric Switches, the first book written by Brocade engineers for designing, building, and maintaining SANs with Brocade switches.

ISBN: 1-928994-30-X

Price: $79.95 US, $123.95 CAN

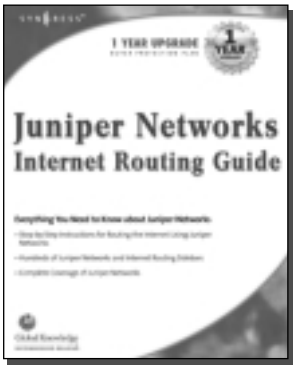
AVAILABLE NOVEMBER 2001
ORDER at
www.syngress.com

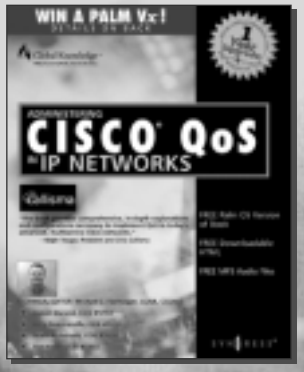
Juniper Networks Internet Routing Guide

Juniper Networks Internet Routing Guide is a comprehensive, hands-on guide on how to install, configure, and troubleshoot Juniper's core router family. There are many similarities between Cisco and Juniper router configuration, and this book makes regular reference through special "How Cisco Does it" sidebars, saving the reader valuable time by making easy comparisons. This will be the book for any engineer migrating to Juniper routers and will immediately meet the demands of network engineers responsible for the installation and configuration of Juniper Routers.

ISBN: 1-928994-76-8

Price: $69.95 US, $108.95 CAN





AVAILABLE NOW
ORDER at
www.syngress.com

Administering Cisco QoS in IP Networks

Administering Cisco QoS in IP Networks discusses IP Quality of Service (QoS) and how it applies to Enterprise and Internet Service Provider (ISP) environments. It reviews routing protocols and QoS mechanisms available today on Cisco network devices. This book will provide you with examples and exercises for a hands-on experience designed to give you the background to implement these capabilities in your network.

ISBN: 1-928994-21-0

Price: $59.95 US, $92.95 CAN

solutions@syngress.com

SYNGRESS®