Cisco Video and TelePresence Architecture Design Guide

March 30, 2012

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
http://www.cisco.com
Tel: 408 526-4000
800 553-NETS (6387)
Fax: 408 527-0883

Text Part Number: OL-27011-01

Any Internet Protocol (IP) addresses and phone numbers used in this document are not intended to be actual addresses and phone numbers. Any examples, command display output, network topology diagrams, and other figures included in the document are shown for illustrative purposes only. Any use of actual IP addresses or phone numbers in illustrative content is unintentional and coincidental.

Cisco Video and TelePresence Architecture Design Guide

CONTENTS

Preface

Revised: March 30, 2012, OL-27011-01

This document describes some of the concepts related to video communications, and it presents considerations and guidelines for designing a video-enabled network.

This document should be used in conjunction with other documentation available at the following locations:

- For Cisco Collaboration and Unified Communications System design documents:

  http://www.cisco.com/go/ucsrnd

- For other Cisco design guides:

  http://www.cisco.com/go/designzone

- For Cisco TelePresence and Video product documentation:

  http://www.cisco.com

Revision History

This document may be updated at any time without notice. You can obtain the latest version of this document online at:

http://www.cisco.com/go/ucsrnd

Visit this Cisco.com website periodically and check for documentation updates by comparing the revision date of your copy with the revision date of the online document.

The following table lists the revision history for this document.

Revision Date	Document Part Number	Comments
March 30, 2012	OL-27011-01	Initial version of this document.

Obtaining Documentation and Submitting a Service Request

For information on obtaining documentation, submitting a service request, and gathering additional information, see the monthly *What's New in Cisco Product Documentation*, which also lists all new and revised Cisco technical documentation, at:

http://www.cisco.com/en/US/docs/general/whatsnew/whatsnew.html

Subscribe to the *What's New in Cisco Product Documentation* as an RSS feed and set content to be delivered directly to your desktop using a reader application. The RSS feeds are a free service. Cisco currently supports RSS Version 2.0.

Conventions

This document uses the following conventions:

Convention	Indication
bold font	Commands and keywords and user-entered text appear in **bold** font.
italic font	Document titles, new or emphasized terms, and arguments for which you supply values are in *italic* font.
[ ]	Elements in square brackets are optional.
{x | y | z }	Required alternative keywords are grouped in braces and separated by vertical bars.
[x | y | z]	Optional alternative keywords are grouped in brackets and separated by vertical bars.
string	A nonquoted set of characters. Do not use quotation marks around the string or the string will include the quotation marks.
`courier` font	Terminal sessions and information the system displays appear in `courier` font.
< >	Nonprinting characters such as passwords are in angle brackets.
[ ]	Default responses to system prompts are in square brackets.
!, #	An exclamation point (!) or a pound sign (#) at the beginning of a line of code indicates a comment line.

Note Means *reader take note*.

Tip Means *the following information will help you solve a problem.*

Caution Means *reader be careful*. In this situation, you might perform an action that could result in equipment damage or loss of data.

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

Warning **Means *reader be warned*. In this situation, you might perform an action that could result in bodily injury.**

CHAPTER 1

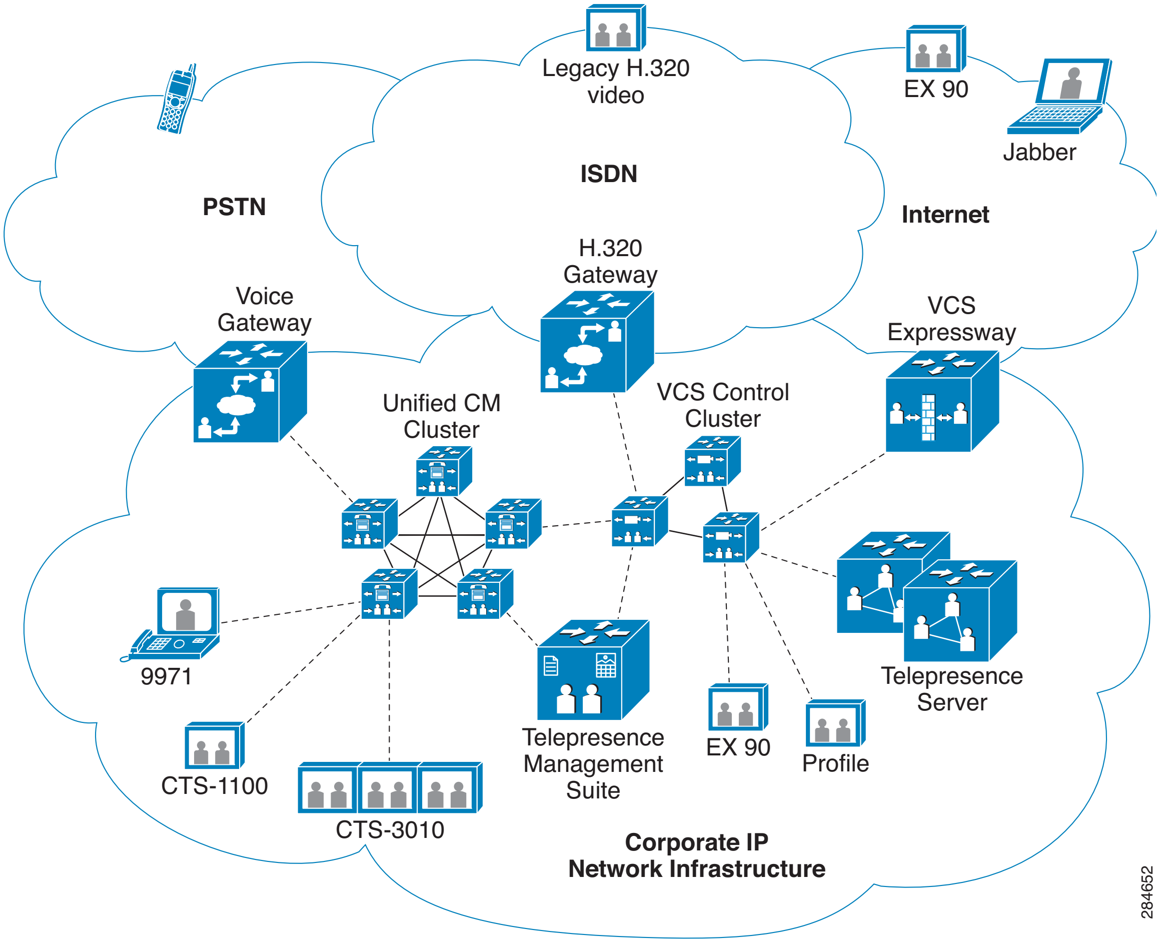
Introduction

Revised: March 30, 2012, OL-27011-01

This document focuses on the two-way interactive video solutions available from Cisco, including Cisco TelePresence and Cisco Unified Communications, and it provides an explanation of the overall solution, technology components, and deployment considerations. With the TelePresence and Unified Communications product families, Cisco offers a wide range of video solutions from interactive video applications for large boardrooms down to mobile users. Cisco also offers a comprehensive set of one-way video applications such as streaming video, digital signage, video surveillance, and even media transformation, which are not covered in this document.

Each solution, Cisco TelePresence or Cisco Unified Communications, can be deployed as a standalone solution or together as an integrated solution. Figure 1-1 illustrates an example of a video architecture that supports both TelePresence and Unified Communications video endpoints. This specific example also shows access to the PSTN for voice calls, ISDN for legacy video, and the Internet-based video device.

Figure 1-1 Cisco TelePresence and Unified Communications Video Architecture



This architecture incorporates the endpoints and infrastructure components listed in Table 1-1 and Table 1-2, respectively.

Table 1-1 Current Cisco TelePresence and Unified Communications Video Endpoints

Category	Endpoint
TelePresence – Immersive	TX9000 Series CTS 3000 Series CTS T Series TX1300 Series
TelePresence – Multipurpose	CTS MX Series CTS Profile MXP Series CTS Profile Series
TelePresence – Desktop	CTS EX Series CTS MXP Series
TelePresence –Office	CTS 1100 CTS 500
TelePresence – Personal	Cisco Jabber Video for TelePresence (Movi)
TelePresence – Video phone	CTS E20
Unified Communications – Video phone	Cisco Unified IP Phones 9900 Series
Unified Communications – Desktop	Cisco Unified Personal Communicator Cisco Jabber
Unified Communications – Tablet	Cius

Table 1-2 Cisco Video Infrastructure Products

Purpose	Product Name	Product Category	Description
Call Control	Cisco Unified Communications Manager	Unified Communications and TelePresence	Call control for Unified Communications and TelePresence devices
	Cisco TelePresence Video Communications Server	Control	Call control for TelePresence and videoconferencing devices
Conferencing	Cisco Integrated Services Router G2 Conferencing Services	Unified Communications	Multipoint conferencing for all video endpoints except three-screen immersive
	Cisco TelePresence Server	TelePresence	Multipoint Control Unit for all video endpoints, including three-screen immersive
	Cisco TelePresence Conductor	TelePresence	Policy server for multipoint device management
	Cisco TelePresence Multipoint Control Unit	TelePresence	Multipoint Control Unit for all video endpoints except three-screen immersive
	Cisco TelePresence Multipoint Switch	TelePresence	Multipoint switch for CTS, EX, and Profile series video endpoints
Gateways	ISDN Gateway	TelePresence	Video gateway allowing connectivity from H.323 and SIP video endpoints to ISDN H.320 endpoints
	Advanced Media Gateway	TelePresence	Gateway allowing connectivity from standard H.323 and SIP video endpoints to Microsoft Lync and Office Communicator devices
	Cisco Telepresence Video Communications Server Expressway	TelePresence	Gateway that provides secure communications between SIP and H.323 video endpoints across the internet
	Cisco Unified Border Element	TelePresence	Gateway that provides a secure demarcation between IP networks
	Cisco Intercompany Media Engine	Unified Communications	Gateway that provides intercompany connectivity when used with Unified CM and ASA firewalls
Recording and Streaming	Cisco TelePresence Content Server	TelePresence	Recording and streaming for all video endpoints
	Cisco TelePresence Recording Server	TelePresence	Recording server for CTS series video endpoints
Management	Cisco TelePresence Manager	TelePresence	Scheduling and management platform for Cisco and third-party video endpoints
	Cisco TelePresence Management Suite	TelePresence	Scheduling and management platform for Cisco and third-party video endpoints
	Cisco Prime Collaboration Manager	TelePresence	Network and endpoint management for media flows

Addressing complex customer requirements is possible due in part to the large number of Cisco TelePresence and Unified Communications video endpoints and infrastructure components. However, a large number of options can make choosing the right solution difficult.

As you will see throughout this document, Cisco TelePresence and Cisco Unified Communications endpoints and infrastructure components share the same protocols, audio and video codecs, and similar deployment considerations. This document explores deeper into the following areas related to Cisco TelePresence and Cisco Unified Communications:

- Video components

  Video components consist of devices such as video endpoints, call control, conferencing, gateways, and management platforms.

- Basic concepts and terminology of video solutions

  TelePresence and video in general introduce a lot of new terminology and concepts that are not found in other technologies. In just the past few years many new products and features have been introduced with the advancement of video endpoints, conferencing devices, and error concealment.

- Call control protocols

  Call control protocols handle the setup and processing of media flows across the network. A number of video call control protocols are used for transporting interactive video over various network media.

- Quality of Service (QoS) and call admission control

  Interactive video is very sensitive to delay, loss, and jitter. Allowing admission to the network only when bandwidth is available and guaranteeing media flows that meet Service Level Agreements (SLAs) are key factors to any successful video deployment.

- Dial plan

  Dial plans provide call routing between video devices and devices external to an enterprise, such as video intercompany calls over the Internet and the PSTN, as well as PSTN audio-only calls. Dial plans might need special attention, depending on the method used to address endpoints or to support advanced feature sets.

- Deployment scenarios

  There are a number of deployment scenarios available for interactive video deployments. Deployment scenarios are based on a number of factors such as the number and type of endpoints, and they focus on all aspects of call control, video services, and network design.

- Business-to-business (B2B)

  Business-to-business video communications is becoming more important as video continues to be deployed and used by more companies. There are a number of ways to allow business-to-business video communications, depending on the call control platforms and endpoints used in an enterprise.

- Conferencing

  Conferencing allows more than two devices to communicate in a single meeting. There are a number of options for initiating a conference and a number of different platforms available for conferencing video endpoints.

- Security

  Security for video calls is a must for many enterprises, especially those using video for business-to-business communications. There are a number of methods available for encrypting signaling traffic and media, and a number of factors that must be considered when deploying secured video communications.

In addition to this document, a number of design and deployment guides have been written to help users choose the correct architecture. These guides not only help with video architecture, but in many cases they also outline the network requirements to ensure proper handling of video calls across the network. The following design and deployment guides cover the deployment of both Cisco TelePresence and Cisco Unified Communications video:

- Cisco Unified Communications System SRND

  http://www.cisco.com/go/ucsrnd
- Cisco TelePresence Network Systems Design Guide

  http://www.cisco.com/en/US/solutions/ns340/ns414/ns742/ns819/landing_vid_tPresence.html

CHAPTER 2

Video Infrastructure Components

Revised: March 30, 2012, OL-27011-01

Video deployments continue to expand as enterprises focus on decreasing travel, increasing productivity, and expanding video in Unified Communications platforms. As the Unified Communications and TelePresence markets continue to grow and mature, the line between the two markets continues to blur. Both TelePresence and Unified Communications video devices employ many of the same protocols and codecs, providing full integration and the ability to utilize infrastructure devices from both solutions.

Both Cisco TelePresence and Cisco Unified Communications solutions have expanded dramatically with internal development and strategic acquisitions by Cisco.

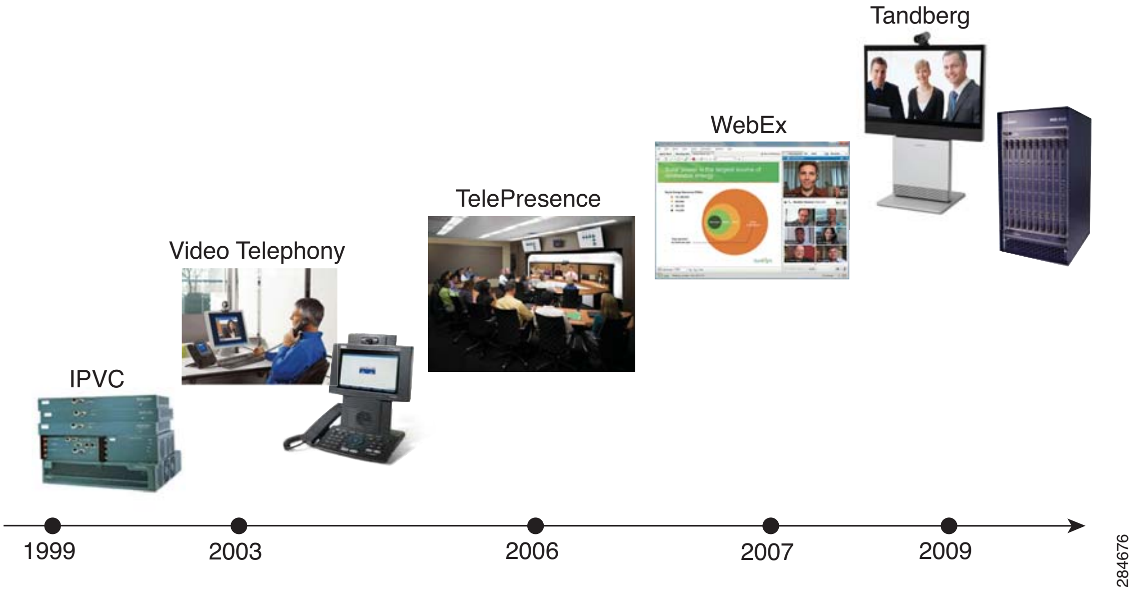
Cisco TelePresence and Unified Communications

Cisco initially entered the two-way interactive video space in late 1999 through an OEM relationship with Radvision. The Radvision relationship brought to market a number of Cisco-labeled videoconferencing infrastructure products, including the Cisco IPVC Multipoint Control Units (MCUs) and IPVC Gateways. A few years later Cisco added video support to its voice call control agent, Cisco Unified Communications Manager (Unified CM), and Cisco also added a number of video telephony products that expanded its video offering to PC-based soft clients and videophones.

In late 2006 Cisco introduced Cisco TelePresence, which added a complete portfolio of high definition (HD) conferencing products, including endpoints, a multipoint switch, TelePresence management, and a recording server. Cisco TelePresence took off and rejuvenated the videoconferencing market, making HD the norm.

With the acquisition of WebEx in 2007, Cisco added more Unified Communications products to its portfolio, including a new video soft client. Finally, in 2009 Cisco acquired Tandberg. At the time of the acquisition Tandberg was the number-one videoconferencing vendor in the market, with the most comprehensive product line in the industry. Combining Cisco TelePresence with the Tandberg portfolio instantly gave Cisco the best-of-breed TelePresence products from the desktop to the boardroom. Figure 2-1 illustrates the progression of interactive video within Cisco's product portfolio.

Figure 2-1 Progression of Cisco Two-Way Interactive Video Products



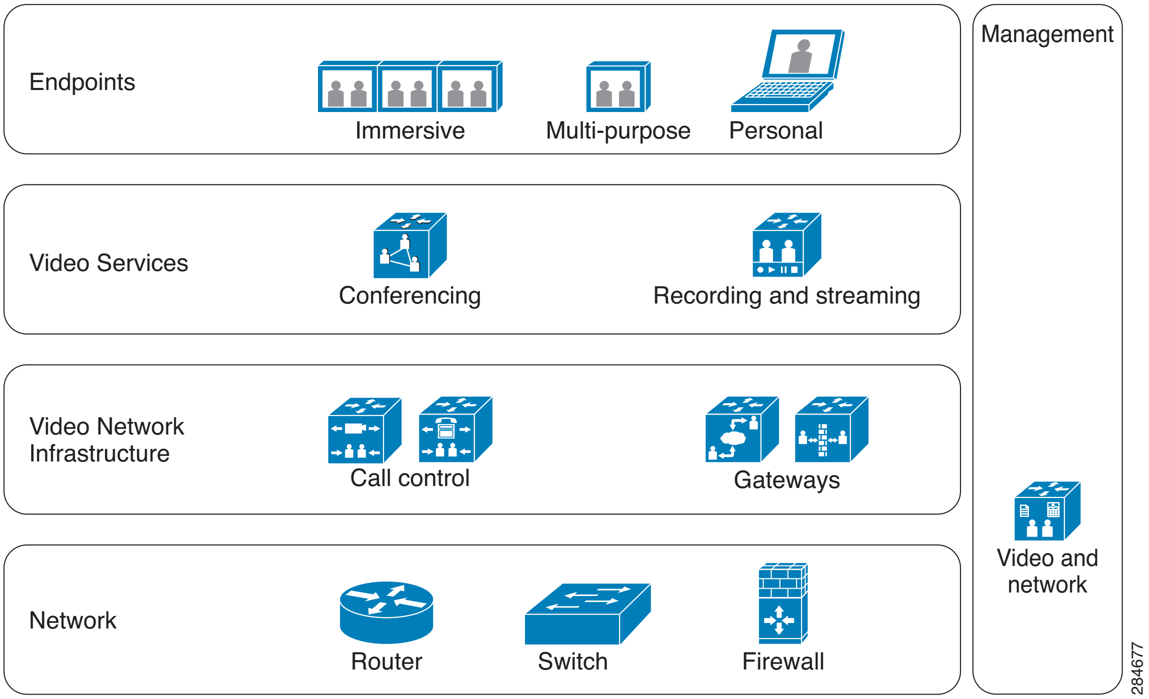
Video Architecture

All interactive video architectures consist of five categories, as shown in Figure 2-2:

- Endpoints, page 2-3
- Video Services, page 2-4
- Video Network Services, page 2-6
- Management, page 2-8
- Network, page 2-9

Each category contains devices that provide a specific function for the video deployment, but devices from all categories are not required or present in all video deployments.

Figure 2-2 Video Architecture



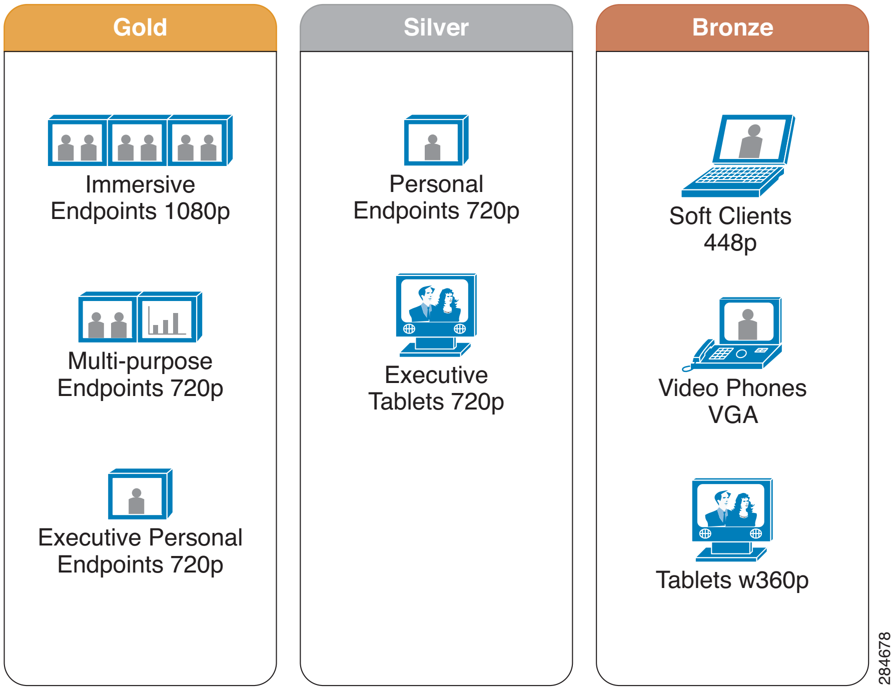
Endpoints

An endpoint consists of a screen, microphone, speakers, and one or more video and audio processing devices called codecs. These components are usually combined into a single unit that can range from a phone with a screen (at the basic end), to a large TV-like device, to an immersive multi-screen system with integrated tables and seating. Cisco provides a large number of video endpoints ranging from video-enabled tablets to multi-screen immersive endpoints.

Different types of video endpoints provide different user experiences and in many cases different feature sets. Each video endpoint supports multiple resolutions, but not all video endpoints support the same set of resolutions. For example, a multi-screen immersive endpoint might support high definition resolutions up to 1080p at 30 frames per second (fps) while a video-enabled tablet might support high definition resolutions up to only 720p at 30 fps. All video endpoints have a core set of features that usually consist of the ability to send and receive live audio and video, and send and receive shared content. Depending on the endpoint type, advanced features such as integrated conferencing or the ability to support additional video and audio sources may be available.

It is important to remember that higher resolutions consumes more network bandwidth. In most deployments, customers choose to limit the higher resolutions based on the endpoint type or user type. Video deployments always start with the type of video endpoints being deployed and the resolutions being supported. It is not uncommon to see customers create multiple classes of video that directly correlate with the supported resolution or user type.   Classes are often based on a minimum level of service provided to each endpoint type or user type, and the classes may contain different maximum resolutions for the same type of endpoint depending on the user. Figure 2-3 shows an example of possible video classes deployed in an enterprise.

Figure 2-3 Example of Possible Video Classes



Video Services

Video services consist of two subcategories:

- Conferencing, page 2-4
- Streaming and Recording, page 2-5

Video services are not required but are an important part of any video deployment. Almost all video deployments use one or more of these services.

Conferencing

Conferencing devices allow three or more video devices to participate in a meeting at the same time. Some conferencing devices can also provide management of conferencing resources, thus allowing conferencing ports to be used more efficiently. Cisco provides conferencing devices that support switching and transcoding.

Switching forwards incoming audio and video without manipulating the video media itself. Switching platforms essentially switch video from one endpoint to another and require all video endpoints in a meeting to send and receive the same resolution. Switching devices offer a cost-effective and scalable solution for video deployments supporting video endpoints with the same resolution sets and not requiring advanced video features such as continuous or active presence.

Transcoding is the encoding and decoding of video media streams between endpoints. Transcoding devices offer support for video endpoints participating in the same meeting with different resolutions, continuous or active presence, and other advanced conferencing features. They allow for maximum conferencing flexibility and feature sets.

Cisco provides multiple video conferencing platforms that support both switching and transcoding. Table 2-1 shows which devices support switching or transcoding.

Table 2-1 Conferencing Platforms for Switching and Transcoding

Conferencing Platform	Switching	Transcoding
Cisco TelePresence Multipoint Switch	Yes	No
Cisco TelePresence Server	No	Yes
Cisco TelePresence MCU 4000 Series and MSE 8000 Series	No	Yes
Cisco Integrated Services Router (ISR) G2	No	Yes

Cisco TelePresence Conductor is a new type of conferencing device that has the capability to manage conferencing ports intelligently. Cisco TelePresence Conductor serves as the front end for all conferencing devices and manages the distribution of conferencing requests. Cisco TelePresence Conductor enables large pools of distributed conferencing resources to be allocated dynamically instead of being limited to conferencing devices with static configurations.

Streaming and Recording

Streaming and recording devices provide the ability to record, replay, and stream important meetings, messages, or updates. Meetings can also be streamed to allow a large number of users to participate in a meeting as view-only participants. Cisco offers one recording and streaming server for TelePresence and Unified Communications video endpoints, and one recording-only server for TelePresence endpoints only, as follows:

- Cisco TelePresence Content Server (TCS)

  The Cisco TCS is available as an appliance or a blade for the Cisco TelePresence Media Services Engine (MSE) chassis. Cisco TCS provides live recording, streaming, and playback of video meetings with content from any TelePresence or Unified Communications video endpoint. Live streams and recordings can be viewed with standard QuickTime, RealPlayer, and Windows Media Players.

- Cisco TelePresence Recording Server (CTRS)

  The CTRS is a server-based platform that provides studio mode and event recording and playback for Cisco TelePresence System 3*x*10, 1300, 1100, and 500 endpoints. Recordings can be viewed in native 1080p or 720p resolution from any Cisco TelePresence System 3*x*10, 1300, 1100, and 500 endpoints. or with QuickTime, RealPlayer or Windows Media Player in CIF format.

Video Network Services

Video network services also consist of two subcategories:

- Call Control, page 2-6
- Gateways, page 2-7

Video network services offer essential services such as call routing and access to external video networks.

Call Control

The main functions of a call control device are endpoint registration, call routing, monitoring, and maintaining connections. Call control platforms also form the base for network dial plans and options for call admission control. Cisco offers two main call control platforms for interactive video: Cisco Unified Communications Manager (Unified CM) and Cisco TelePresence Video Communication Server (VCS). Unified CM has been used for call control and provisioning for some of the largest IP voice deployments in the world. Unified CM is also the call control and provisioning platform for the original Cisco TelePresence and Cisco Unified Communications devices.

Cisco VCS was designed to provide call control for H.323 and Session Initiation Protocol (SIP) video environments with advanced video features to support large-scale deployments. VCS can be deployed in either of the following ways:

- VCS Control — Provides call control for an enterprise video deployment.
- VCS Expressway— Supports Network Address Translation (NAT) and firewall traversal, extending video outside the enterprise for business-to-business communication or internet-based remote workers.

Each call control platform can be deployed independently or as an integrated solution for existing and new customers. Unified CM supports direct registration for all Unified Communications video endpoints and most TelePresence endpoints, while VCS supports most TelePresence endpoints but does not support Unified Communications endpoints. Table 2-2 lists the call control support for each video endpoint type or series.

Table 2-2 Call Control Support Per Video Endpoint

Endpoint	Unified CM	VCS
TX9000 Series	Yes	No
CTS 3000 Series	Yes	No
CTS T Series	No	Yes
TX1300 Series	Yes	No
CTS MX Series	Yes	Yes
CTS Profile MXP Series	Yes	Yes
CTS Profile Series	Yes	Yes
CTS EX Series	Yes	Yes
CTS MXP Series	No	Yes
CTS 1100	Yes	No
CTS 500	Yes	No

Table 2-2 Call Control Support Per Video Endpoint (continued)

Endpoint	Unified CM	VCS
Cisco Jabber Video for TelePresence (Movi)	No	Yes
CTS E20	Yes	Yes
9900 Series	Yes	No
Cisco Unified Personal Communicator	Yes	No
Cisco Jabber	Yes	No
Cius	Yes	No

Gateways

Video gateways provide access from one network to another. Cisco provides the following video gateways:

- ISDN gateways

  ISDN gateways provide TelePresence and Unified Communications video endpoints with connectivity to legacy H.320 video endpoints. ISDN gateways are often referred to as H.320 gateways.

- Advanced media gateways

  Advanced media gateways provide communication between Microsoft Office Communications Server 2007 R2 or Microsoft Lync Server users and standards-based TelePresence and video conferencing devices.

- IP-to-IP gateways

  Cisco offers the following IP-to-IP gateways that provide business-to-business (B2B) connectivity as well as support for Internet connectivity for video endpoints:

  - Cisco VCS Expressway

    VCS Expressway is an appliance that works in conjunction with the VCS Control to provide firewall traversal using H.460.18, Assent, or SIP protocols. It supports Traversal Using Relay NAT (TURN) servers. VCS Expressway also provides endpoint registration and signal and media interworking for SIP and H.323 video devices across the public Internet.

  - Cisco Unified Border Element

    Cisco Unified Border Element is available on a number of Cisco router platforms and provides a network-to-network demarcation point for signaling interworking, media interworking, address and port translations, billing, security, quality of service (QoS), and bandwidth management. Service provider TelePresence Exchanges often implement Cisco Unified Border Element as an IP-to-IP gateway because it provides a demarcation point and adds security between customer networks.

  - Cisco Intercompany Media Engine (IME)

    The Cisco IME is a server-based platform that provides business-to-business connectivity when used in conjunction with the Cisco ASA 5500 Series Adaptive Security Appliances and Cisco Unified CM 8.*x* or later releases.

Management

Video management platforms perform multiple functions such as scheduling, video endpoint and infrastructure monitoring, and in some cases provisioning or tracing of media flows across the network. Cisco provides three main management platforms for TelePresence and Unified Communications video:

- Cisco TelePresence Management Suite, page 2-8
- Cisco TelePresence Manager, page 2-8
- Cisco Prime Collaboration Manager, page 2-8

Cisco TelePresence Management Suite

The Cisco TelePresence Management Suite (TMS) is delivered as a management appliance or software that can be loaded on a server. Cisco TMS provides one-button-to-push (OBTP) call launching, scheduling, monitoring, and provisioning for TelePresence endpoints registered with the VCS. The TMS also provides OBTP, scheduling, and statistics for TelePresence endpoints registered to Unified CM. TMS also extends scheduling and some monitoring features to third-party telepresence and video endpoints such as Polycom and LifeSize.

The TMS can be integrated with enterprise calendaring systems such as Microsoft Exchange for scheduling with tools such as Microsoft Outlook. TMS also has a built-in web scheduling interface that enables users to schedule meetings directly through TMS.

Cisco TelePresence Manager

Cisco TelePresence Manager is a server-based platform that was originally developed to provide OBTP call launching, scheduling, and management for Cisco TelePresence endpoints. Cisco TelePresence Manager also performs scheduling for telepresence endpoints, including third-party endpoints that are not registered to Unified CM.

Cisco TelePresence Manager can be integrated with enterprise calendaring systems such as Microsoft Exchange for scheduling with tools such as Microsoft Outlook. Unlike TMS, Cisco TelePresence Manager does not have built-in web-based scheduling.

Cisco Prime Collaboration Manager

Cisco Prime Collaboration Manager is a server-based network management platform that allows real-time monitoring and analysis of medial flows from Cisco Medianet-enabled devices. Cisco Medianet devices are routers and switches that support Cisco Mediatrace, which maps the path media takes through the network and which can be used only with endpoints that contain the Cisco Media Services Interface (MSI). The MSI is a software component that is embedded in video endpoints and collaboration applications, and it provides advanced features such as auto-configuration of network ports and initiation of Mediatrace. Cisco Prime Collaboration Manager also supplies historical reporting and a view of utilization and problem trends as well as critical outages.

Network

A properly designed network is a key component of any video design. Using existing network protocols, features, and tools simplifies video deployments and helps ensure a successful deployment. Interactive video devices are sensitive to loss, so it is imperative to keep loss to a minimum. Identifying video traffic on the network and ensuring end-to-end Quality of Service (QoS) provides the most predictable video experience.

With the use of protocols such as Cisco Discovery Protocol (CDP), which is a Cisco proprietary Data Link Layer protocol used to share information between network devices, video endpoints can be identified automatically, thus allowing their QoS markings to be trusted, traffic to be placed in the appropriate Virtual Local Area Network (VLAN), and packets to be queued appropriately. Additionally, VLANs can be used to insulate video traffic from other network traffic and to provide additional security.

Video-aware networks allow for real-time traffic analysis for troubleshooting network issues in real time. Tracking video flows through the network and identifying the exact point in the network where loss is occurring is essential in today's networks where video flows between two endpoints can take different paths across the network, depending on network conditions. Using Cisco Medianet-enabled devices not only allows for real-time traffic analysis but also provides utilization data that helps avoid network oversubscription.

CHAPTER 3

Basic Video Concepts

Revised: March 30, 2012, OL-27011-01

This chapter explains some of the fundamental concepts and terminology encountered in video solutions.

Common Terminology in IP Video Solutions

The vocabulary of IP video solutions encompasses a wide range of concepts and terms, from the video stream formation to how and what devices put the video stream in the wire. This section covers the most common concepts and terms, explains how they relate to the IP video technologies, and attempts to de-mystify them.

Video Frame

A video is an action sequence formed by a series of images, and each image in the series succeeds the previous one in the timeline of the action sequence to be displayed. These still images are called video frames. The smaller the time difference is between each video frame, the higher the refresh rate is and the more naturally movement is represented in the video. This is because there are more video frames generated per second to be displayed as part of the action sequence, therefore changes in the image between frames are slighter and movement appears smoother.

Compression in IP Video Solutions

Compression of IP video, as the term implies, is a process by which the overall size of the video information is compacted. Unlike the audio data in an IP telephony stream, which is very light weight, video data is inherently large in size but irregular in its stream flow. The flow irregularity is due to the fact that in video there are portions of the information that remain constant (for example, backgrounds) and portions that are in motion (for example, people). Additionally, motion is not always constant or from the same object size, therefore transmission of real-time video requires complex mechanisms to reduce its size and irregularity. Compression reduces the video size so that it can be transmitted more easily. The primary compression methods for IP video are:

- Lossless, page 3-2
- Lossy, page 3-2

Both of these video compression methods can use the following techniques:

- Intra-Frame, page 3-2
- Inter-Frame, page 3-2

Lossless

Lossless IP video compression produces, on the decompression end, an exact copy of the image that was originally submitted at the input of the compression process. Lossless compression is achieved by removing statistically redundant information so that the receiving end can reconstruct the perfect video signal. That is, there is no intentional loss or pruning of video information that occurs as part of the compression process. Lossless compression is used mostly for archiving purposes because of its inherent ability to preserve everything about the original image. Lossless video compression is rarely used in IP video solutions because it creates a large quantity of information that poses difficulties for streaming.

Lossy

Lossy video compression is more common in IP video than its lossless counterpart. Lossy video compression is based on the premise that not all the video information is relevant or capable of being perceived by the viewer, therefore some video information is intentionally discarded during the compression process. An example of this "irrelevant" video information is noise in the case of video that has undergone analogue to digital conversion. Lossy video compression achieves a very significant decrease in payload size while maintaining a very high presentation quality, thus making it the compression method of choice for IP video solutions. It is important to note that video compression is always a trade-off between video size and quality. Another trade-off is the frame duration or frame rate, which is measured in frames per second (fps). For example, an image with resolution of 720p at 60 fps is more appealing than an image with 1080p at 30 fps because of the savings of roughly 10% bandwidth and better perception of motion.

Intra-Frame

The intra-frame technique consists of compressing the contents of a single video frame at a time, without considering previous or succeeding video frames. Because every video frame is compressed individually, no previous or succeeding compressed video frames are needed to decompress a specific compressed video frame; it is literally as if every compressed video frame is a key frame.

Intra-frame compression alone does not offer many advantages for video streaming or video conferencing because the compression ratio is not as high as with inter-frame techniques. Therefore, intra-frame compression is always used in conjunction with the inter-frame compression technique in video conferencing.

Inter-Frame

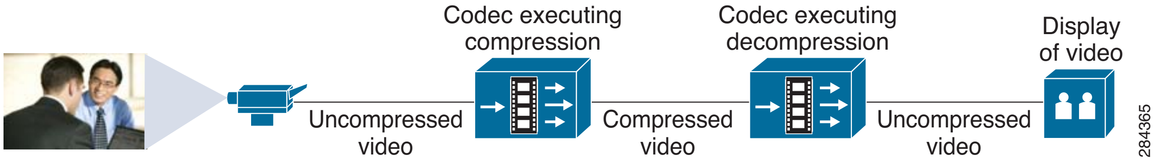
Unlike the intra-frame technique, the inter-frame technique uses information from the preceding video frames to execute the compression. Certain video formats (for example, Advance Video Coding formats such as H.264) that implement the inter-frame technique also use information from the succeeding video frames to perform the compression.

The inter-frame technique relies on the fact that parts of the images to be compressed in video are sometimes not in motion, thus providing the opportunity for the compressor to send only the differences (what has moved) instead of the entire video frame information. It is important to note that this technique relies on the concept of a key frame. The key frame is the initial video frame used for reference in the compression operation. Therefore, the arrival of the key frame at the decoder becomes critical for the decompression operation to be successful. For this reason, video formats that employ the inter-frame technique usually implement recovery mechanisms. Because the key frame is used as a reference in the inter-frame compression technique, its compressed contents do not depend on any preceding or succeeding frames.

Codecs in IP Video Solutions

The term codec stands for COmpressor-DECompressor (or COder-DECoder). Video codecs (such as the Cisco TelePresence System or C-Series codecs) are the hardware or software means by which video is encoded and decoded. In addition, the term *codec* is often used to describe the video formats. A video codec may be able to implement one or more video formats, and these formats may implement lossless or lossy compression methods using either the intra-frame or inter-frame compression technique. In an IP video solution, almost all IP video endpoints integrate a codec as part of their basic functions. As discussed in the section on Compression in IP Video Solutions, page 3-1, compression is necessary because of the large size of video data to be transmitted in a session. Figure 3-1 shows a codec performing compression. The codec helps reduce the video stream size and irregularity by applying a compression operation on it.

Figure 3-1 A Codec Executing Video Compression



Video Compression Formats

As stated in the section on Codecs in IP Video Solutions, page 3-3, video formats are commonly referred to as codecs and the terms are used interchangeably. Video formats are specifications that state how compression or encoding of video takes place using a given technique. For instance, H.264 is a widely used video format that employs the lossy compression method. Video formats are implemented by the codecs employed in the video endpoints to encode the video. IP video endpoints must negotiate and agree on the video format to be used during a call. Although some video formats might implement the same method and technique, they do not necessarily offer the same advantages. How the video format implements the method and technique determines its strengths and advantages.

Generally speaking, video formats are established by the International Telecommunication Union (ITU) Telecommunication Standardization Sector (ITU-T) or by the International Organization for Standardization (ISO) in conjunction with the International Electrotechnical Commission (IEC). Two of the three most common video formats used in IP video solutions (H.261 and H.263) were established by ITU-T, while the third (H.264) is a joint effort of ITU-T, ISO, and IEC (the Moving Picture Experts Group, or MPEG). Table 3-1 compares the features and characteristics of these formats.

Table 3-1 Comparison of Video Compression Formats

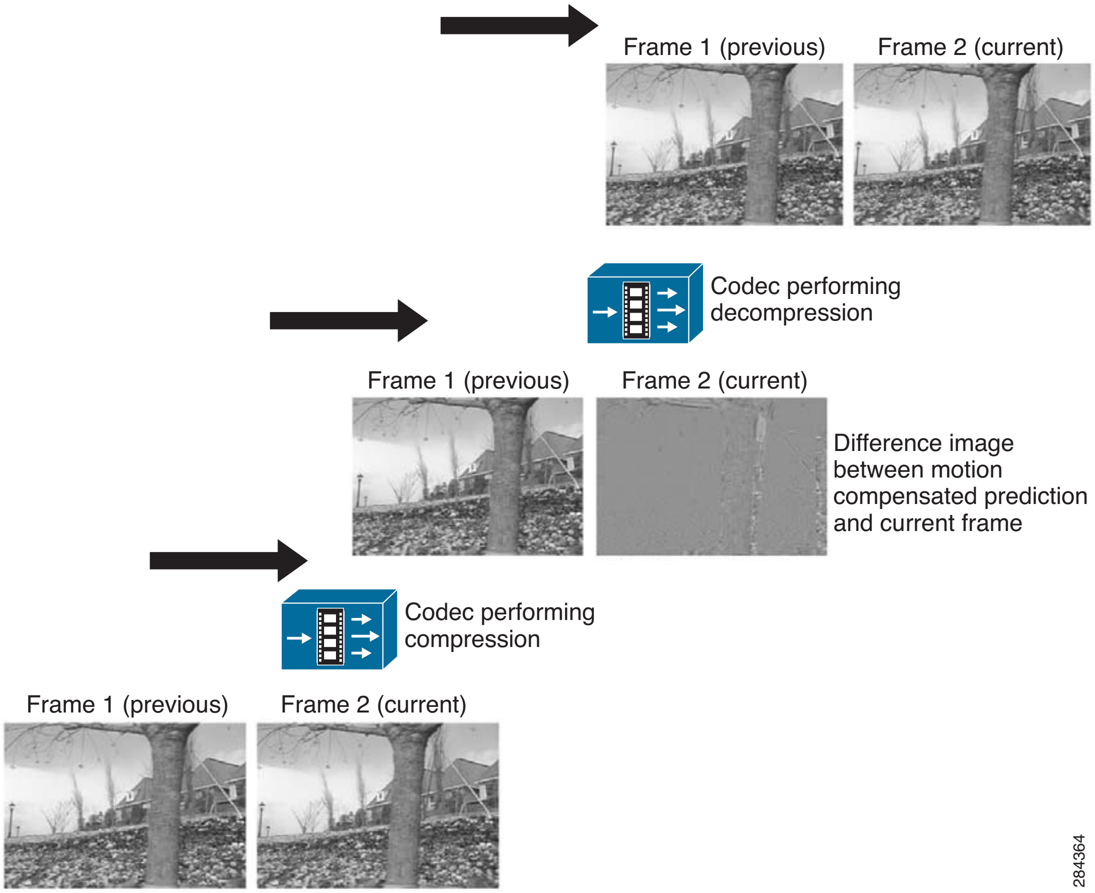
Feature	H.261	H.263	H.264
Bandwidth efficiency	Low	Medium	High
HD support	No	No	Yes
Compressed video frames supported	I-frame, P-frame	I-frame, B-frame, P-frame	I-frame, B-frame, P-frame
Compression and media resiliency features	Error feedback mechanism	Error feedback mechanism Optimized Virtual Channel Link (VLC) tables Four optional negotiation modes (Annex D, E, F, and G)	Error feedback mechanism Enhanced motion estimation Improved entropy coding Intra-prediction coding for I-frames 4x4 Display Channel Table (DCT) Network Abstraction Layer Gradual Decoder Refresh (GDR) frame Long-Term Reference Picture (LTRP) frame

Currently most Cisco IP Video endpoints utilize H.264 as their default video compression format.

Compressed Video Frames

The compressed video frames are the result of the compression operation (using intra-frame or inter-frame techniques and using lossy or lossless methods), and they are used instead of the regular (uncompressed) video frames to reduce the overall size of the video information to be transmitted. Figure 3-2 depicts a video frame being compressed by a codec, and the resulting compressed video frame in this example is an I-frame.

Figure 3-2 Compression at Work


Frame 1 (previous)
Frame 2 (current)
Codec performing decompression
Frame 1 (previous)
Frame 2 (current)
Difference image between motion compensated prediction and current frame
Codec performing compression
Frame 1 (previous)
Frame 2 (current)
284364

There are many different kinds of compressed video frames used in IP video solutions, but the main types are:

- I-Frame, page 3-5
- P-Frame, page 3-6
- B-Frame, page 3-6

I-Frame

I-frames rely only on their own internal data and they enable decompression or decoding at the start of the sequence. I-frames are compressed using the intra-frame method. I-frames are also known as key frames because their content is independent of any other frames and they can be used as a reference for other frames. As discussed in the inter-frame compression method, a key frame or initial frame is used at the beginning of the sequence of pictures to be compressed. Instant Decoder Refresh (IDR) frames, Gradual Decoder Refresh (GDR) frames, and Long-Term Reference Picture (LTRP) frames are well known I-frames. The main difference between IDR and GDR frames is that a GDR frame can be divided into smaller frames and sent at smaller time intervals whereas an IDR frame is sent in one packet. The purpose of using GDR frames is to avoid a significant surge in the data rate that occurs with IDR frames and to provide a better experience of video quality for the end user. For example, a GDR implementation can send 10 individual sections of a complete frame, and each of these sections is itself an IDR encoded video picture. Because only 1/10 of the frame is gradually changing over a 10-frame window, the user perception of the video quality is generally very good.

LTRP frames, on the other hand, are part of the media resilience provisions that some codecs implement. Inevitable network loss and errors of compressed video cause visual errors at the decoder. The errors would spread across subsequent P-frames. An obvious way to avoid this problem is to have the decoder request an I-frame from the encoder to flush out the errors (error feedback mechanism). However, a better way is to employ different reference frames (older, long-term frames). The feedback mechanism, in conjunction with the known good LTRP frame, helps to repair the lost video data (for example, slices) and to phase out the bad LTRP frame. In codecs that support this implementation, the LTRP frame is the last I-frame that arrived at the codec (using either IDR or GDR) The receiving codec then stores this frame as the LTRP frame. When a new I-frame arrives, it is promoted to LTRP, and so on. If an I-frame is lost during transmission, the receiving codec attempts to use the LTRP frame to recover.

I-frames are compressed using the intra-frame technique, and this has a direct impact in the bandwidth consumption of the video stream. The more frequently I-frames are used, the more bandwidth is required.

P-Frame

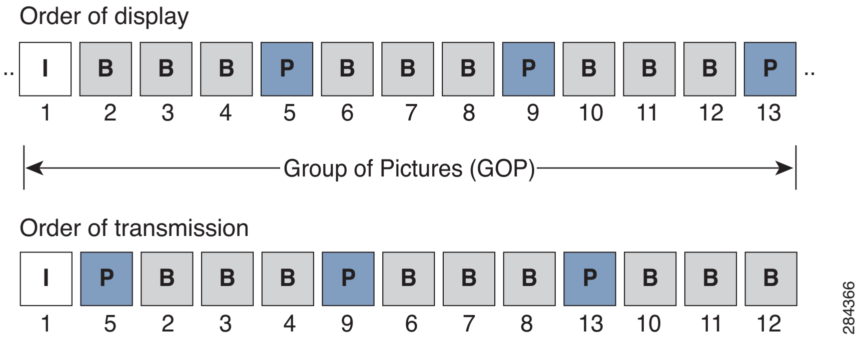
Predictive frames (P-frames) are more compressible that I-frames. P-frames are compressed using the inter-frame encoding technique. P-frames follow I-frames in the video stream, and they store only the video information that has changed from the preceding I-frame. As mentioned in the discussion of inter-frame compression, correct decoding of a P-frame is possible only in combination with the most recent I-frame (key frame).

B-Frame

Although the use of P-frames increase the compression significantly, bidirectional predictive frames (B-frames) make the overall compression more efficient. B-frames reference the previous I-frames and subsequent P-frames, and they contain only the image differences between them. However, not all codecs have the ability to implement B-frames (assuming that the video format utilized in the call supports B-frames) because the codec needs twice as much memory to provide enough buffer between the two anchor frames. B-frames also add some delay that is intrinsic to the logic of the B-frame implementation.

Figure 3-3 depicts the order of the various compressed video frames in a video stream. In this example the codec is able to implement I-frames, P-frames, and B-frames.

Figure 3-3 Order of Video Display



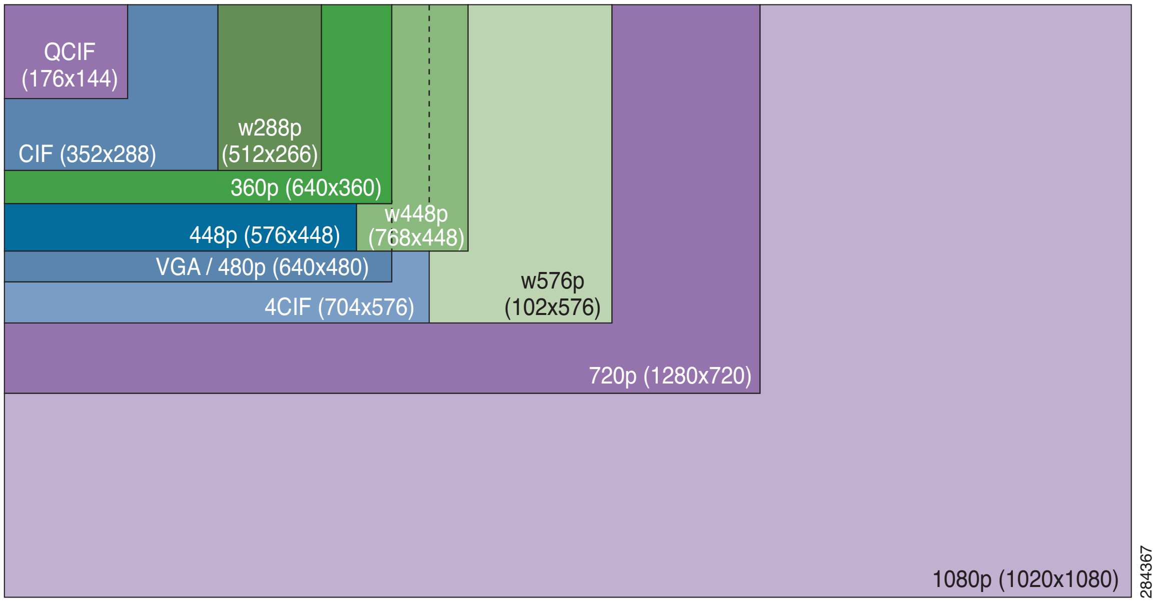
Resolution Format in IP Video Solutions

In simplest terms, the resolution format is the image size. However, it is important to note that most video endpoints today have the ability to scale the image to fit the screen where the video is displayed. Although this is necessary for the video to be visible from afar, it causes the image to be less crisp.

The video resolution format is formally defined as the scanning method used in combination with the distinct number of pixels being presented for display. The following list and Figure 3-4 depict some common video resolution formats in IP video solutions.

- CIF — Common Intermediate Format
- QCIF — Quarter CIF
- 360p — 360 vertical, progressive scan
- 480p — 480 vertical, progressive scan
- 720p — 720 vertical, progressive scan
- 1080i — 1080 vertical, interlaced video
- 1080p — 1080 vertical, progressive scan

Figure 3-4 Popular Video Resolutions



Evolution of Cisco IP Video Solutions

IP video has largely replaced other video conferencing methods. However, interconnection between methods is sometimes necessary, therefore a basic understanding of the other methods is useful. This section briefly highlights the evolution of video conferencing solutions, from video over ISDN media to the newer cloud-hosted video solutions, and the interoperability between those solutions. The following video solutions and their interoperability are covered in this section:

- Video over ISDN, page 3-8
- IP Video Telephony, page 3-9
- Desktop Video Conferencing, page 3-10
- Immersive Video Conferencing, page 3-11
- Cloud-Hosted Video Solutions, page 3-12
- Interoperability, page 3-12

This section does not cover these solutions in strict chronological order; some of the solutions overlap each other or were developed around the same time.

Video over ISDN

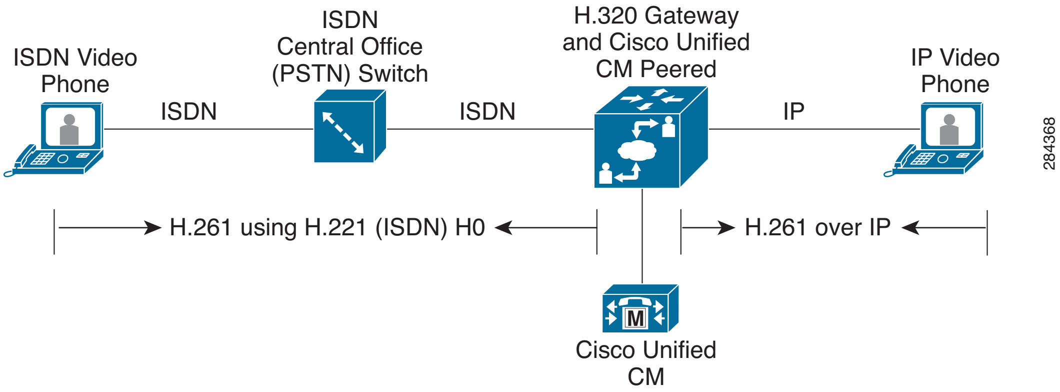
Video conferencing was not widely used until the creation of the Integrated Services Digital Network (ISDN) standard. Therefore, ISDN is often seen as the first technology catalyst that helped to spread video conferencing. As video conferencing matured, new solutions emerged that offered better interoperability, resiliency, and video quality. As Cisco entered the video conferencing market, it became apparent that the ISDN video terminals would need to interoperate with the emerging technologies. To connect the new IP video networks with existent ISDN video terminals, Cisco IP Video Solutions integrated the Cisco Unified Videoconferencing 3500 Series products as H.320 gateways. Since then, Cisco has incorporated a variety of H.320 devices into its portfolio to support the video ISDN space. These H.320 devices peer to a gatekeeper, Cisco Unified Communications Manager (Unified CM), or Cisco TelePresence Video Communication Server (VCS) to provide IP video endpoints with access to ISDN video endpoints residing on the other side of the PSTN cloud.

The H.320 standard defines multimedia (H.221 for video in our case of interest) in ISDN. H.320 originally defined H.261 or H.263 as the video formats to be used when video is used in conjunction with ISDN, and the last update to the standard added H.264. H.221 defines four modes of transmission: Px64 kbps, H0 (384 kbps), H11 (1536 kbps), and H12 (1920 kbps). After the video is encoded, the selected video format (for example, H.261) is multiplexed using the H.221 standard.

ISDN is called a narrow-band visual telephone system because the video resolution formats it supports are very limited in image size. ISDN supports QCIF, CIF, 4CIF, and 16CIF as video resolution formats.

A distinctive characteristic of this kind of solution is its dependency on a supporting ISDN service provider, which remains permanently engaged so that the call can work between the different ISDN terminals, as depicted in Figure 3-5.

Figure 3-5 Image 4. Video over ISDN and Protocols Used

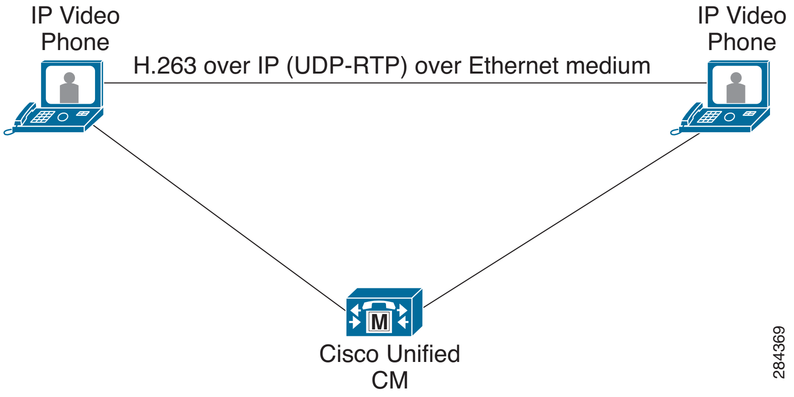

ISDN Video Phone
ISDN
ISDN Central Office (PSTN) Switch
ISDN
H.320 Gateway and Cisco Unified CM Peered
IP
IP Video Phone
H.261 using H.221 (ISDN) H0
H.261 over IP
Cisco Unified CM
284368

IP Video Telephony

While video over ISDN was the first video conferencing technology deployed in practice, IP video telephony brought video conferencing to the enterprise on a much larger scale. IP video telephony enables video in the enterprise through a variety of approaches. It can enable video on the user's IP phone through a software client running on a PC, and it can incorporate specialized video endpoints and video conference bridges to provide a rich media experience. Unlike video over ISDN, IP video telephony provides better video resolution, resilience, and interoperability.

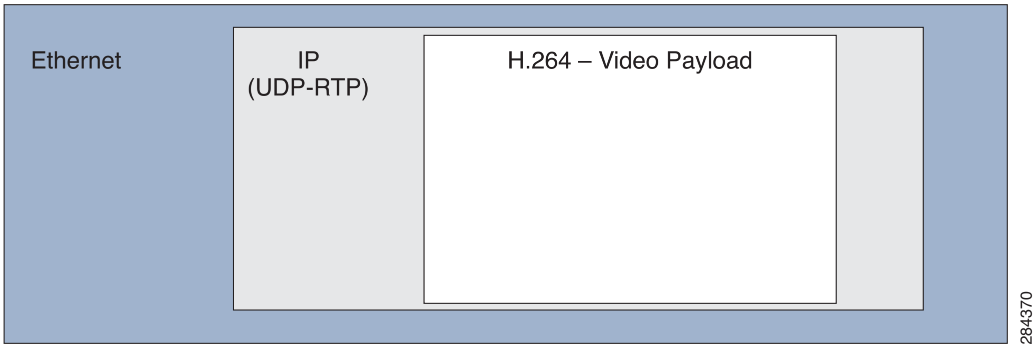
A call control element is an integral part of every IP video telephony solution. This element is responsible for the call routing and, in most cases, interoperability and the handling of special features. In Cisco's first iteration of IP Video Telephony, Cisco Unified Communications Manager (Unified CM) executed call control. Figure 3-6 depicts a sample topology of Cisco IP Video Telephony.

Figure 3-6 IP Video Telephony



As shown in Figure 3-7, IP Video Telephony improves the video resolution by providing transmission flexibility in a variety of physical transport media that are not restricted to 2 Mbps as video over ISDN was (1.54 Mbps in the case of the US). IP (User Datagram Protocol for video) encapsulated packets can be carried over Ethernet, wireless, Multiprotocol Label Switching (MPLS), and so forth. The synergy between the new transmission media (MPLS, Ethernet, optical, and so forth) and IP allows for transmission of larger compressed video frames for increased resolution. Resiliency is boosted by new error recovery techniques implemented by new codecs, while backward compatibility is also maintained.

Figure 3-7 Encapsulation of Compressed Video Frames in IP

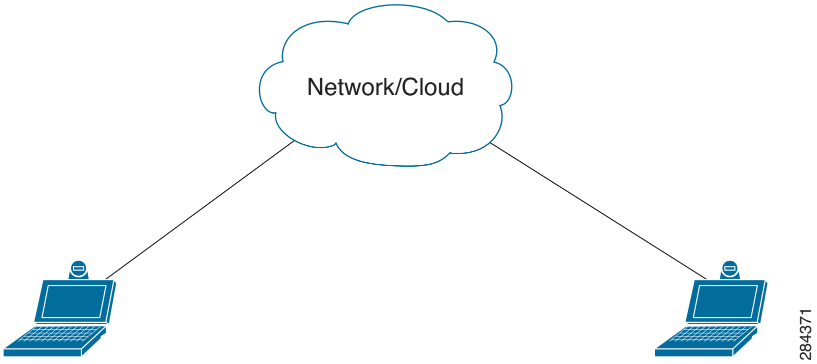


Desktop Video Conferencing

Desktop video conferencing involves the consolidation of IP video as the next generation of communication. Desktop video conferencing started as an add-on to instant messaging programs. In parallel, IP video telephony technology companies realized its benefits and created software video clients that would peer with existent IP telephony deployments. Some technologies leveraged current hardware IP phones and some leveraged software IP phones. Cisco's initial offering for desktop conferencing was Cisco Unified Video Advantage (VT Advantage), a software video client that enables video capabilities in both hardware and software IP phones.

Desktop video software clients use the computer's resources to execute software encoding and decoding of video. The higher the video resolution format and video format complexity, the more computer resources are needed. As faster and better computers became available and more efficient encoding-decoding mechanisms were devised, advanced desktop video conferencing clients became common in the end user space as well. Figure 3-8 shows the typical usage and basic topology of a video software client during a session.

Figure 3-8 Software Video Clients



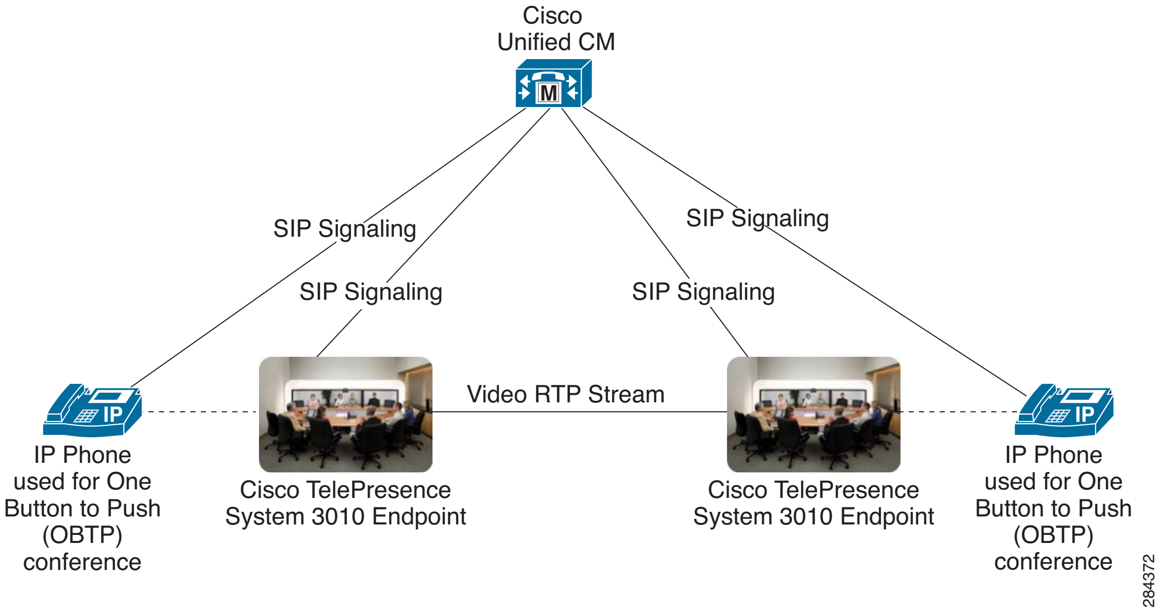
Immersive Video Conferencing

As the quest for new methods of video communications continued, new implementations of IP video solutions were conceived. Life-size video systems, called telepresence, were created as a means of communicating more naturally with remote participants. The first telepresence systems suffered from low adoption rates due to their high cost and dedicated network requirements. In 2006, Cisco entered the immersive video conferencing market, leveraging its vast networking knowledge to create a true converged network telepresence product. Eventually, other immersive video conferencing manufacturers followed Cisco's lead in creating converged network telepresence systems.

Cisco TelePresence shares some aspects in common with regular IP video telephony. Compressed video frames are encapsulated in User Datagram Protocol (UDP), enabling access to the same kind of media IP video telephony uses and providing compatibility with video formats used in IP video telephony. Despite their similarities, though, some elements of Cisco TelePresence differ from IP video telephony. Cisco TelePresence uses high definition cameras and displays, which are specially fitted in the case of large participant rooms. Although the call routing in Cisco TelePresence is still handled by a call agent, the way users interact with the system for call initiation is different than with IP video telephony

Telepresence systems use high definition cameras to capture rich video. After encoding and decoding, this video is displayed on high definition displays to preserve as much of the experience as possible. Additionally, special conditioning of conference rooms for a studio-like setting is available to increase the realism of the meeting. As described earlier, end users interact differently with telepresence systems for meeting initiation. Telepresence systems typically integrate mechanisms to start meetings at the push of a button. In the case of Cisco TelePresence, this session initiation feature is call One Button To Push (OBTP). Figure 3-9 illustrates the flow of media and signaling in a basic point-to-point call for immersive Cisco TelePresence.

Figure 3-9 Immersive Telepresence


Cisco
Unified CM
M
SIP Signaling
SIP Signaling
SIP Signaling
SIP Signaling
Video RTP Stream
IP
IP Phone used for One Button to Push (OBTP) conference
Cisco TelePresence System 3010 Endpoint
Cisco TelePresence System 3010 Endpoint
IP
IP Phone used for One Button to Push (OBTP) conference
284372

Cloud-Hosted Video Solutions

Cloud-hosted video solutions are subscription-based services that provide video communications across the Internet, making enterprise-grade video collaboration both affordable and accessible.

A notable difference between this solution model and the others is that the customer does not front the cost of the IP video infrastructure and acquires only the video endpoints (for example, a Cisco TelePresence System EX90 or a PC). The multiplexing and control of the video endpoints occur off premises, empowering customers to enter into the video collaboration space without a significant investment in the infrastructure. This solution model does require an available internet connection and a subscription from an IP video provide, but the IP video endpoints can be reused if the solutions is migrated to an on-premises model.

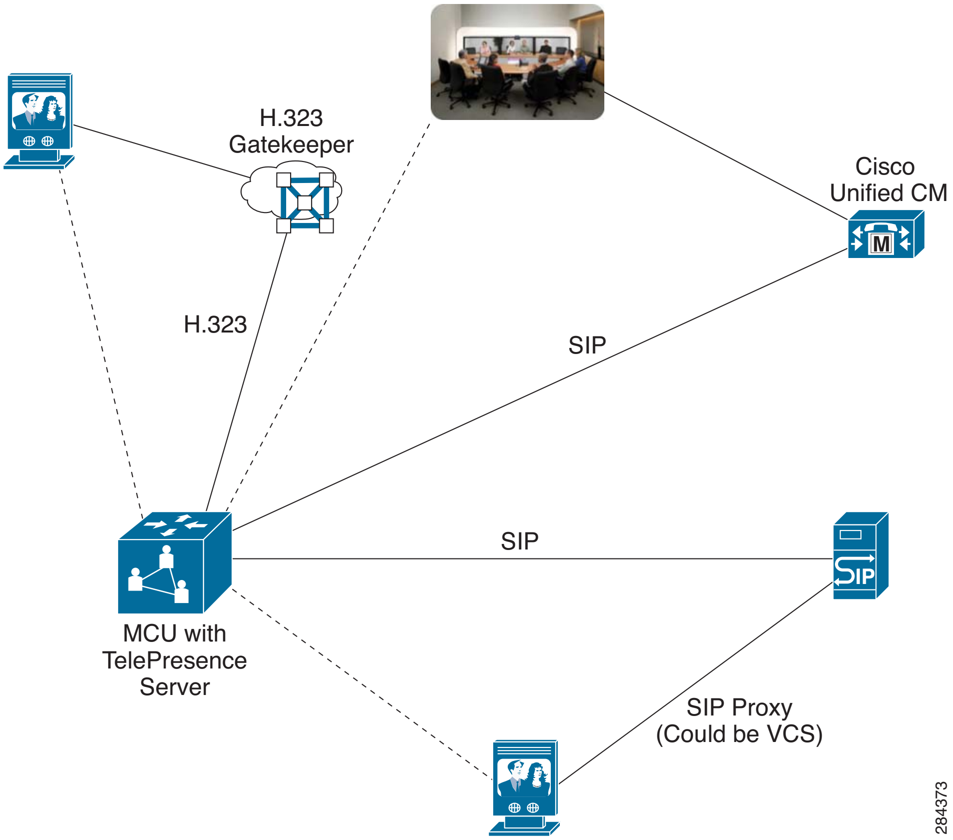
Cloud-hosted video solutions solve the problem of the high cost of an IP video infrastructure by empowering customers to pay as they go for the IP video service. Examples of this type of video solution are Cisco Callway and Cisco WebEx, both of which provide video capability and allow customers to enable video for their users with less administrative overhead and infrastructure investment.

Interoperability

Advancements in technology inevitably create the need for the new technology to interconnect and work with legacy technologies. Interoperability solves the problem of interconnecting different IP video technologies, but interoperability is restricted to features that can be implemented in the target technology. For example, some ISDN terminals are capable of sending text to a participant's screen when on a video call because the ISDN standard provides for text transmission. However, it is not technologically possible to pass this text outside of the ISDN domain (for example, to IP video telephony) because the standards implemented other technologies do not allow for text transmission.

Interoperability is typically achieved using a product or suite of products to provide the edge element between the technology islands. Usually the types of products used (either individually or in combination) to provide interoperability to a video solution include video transcoders, video gateways, and video conference bridges. Figure 3-10 shows a common interoperability scenario, with interoperability provided by a Multipoint Control Unit (MCU).

Figure 3-10 Interoperability in a Video Conferencing System



Legacy Multipoint Control Units

Early Multipoint Control Unit (MCU) architectures offered limited services and capabilities. These legacy MCUs had two main hardware components, a controller blade and a digital signal processor (DSP) blade. The controller blade was aware of only the local DSP assets and therefore it was impossible for it to ascertain the assets of a different MCU to cascade them and use them in a video multipoint call. Furthermore, only certain resolutions were supported, and transrating often was either not supported or came at the sacrifice of high capability.

Although some legacy MCUs added support for high definition video in their later iterations, the majority of the legacy MCUs typically offer support only for standard definition video.

Common Technologies Used in Cisco IP Video Solutions

Although the list of technologies used in IP video solutions is long, this section discusses the technologies currently used Cisco IP Video solutions. With these technologies, Cisco has solved particular problems that otherwise would be left unaddressed. For instance, packet loss, although avoided as much as possible in every deployment, is sometimes inevitable when control over the transmission medium is lacking. Cisco ClearPath helps minimize the impact of packet loss. Telepresence Interoperability Protocol (TIP), on the other hand, addresses several issues, including what video to display when multiple screen systems are talking. This section describes the following technologies:

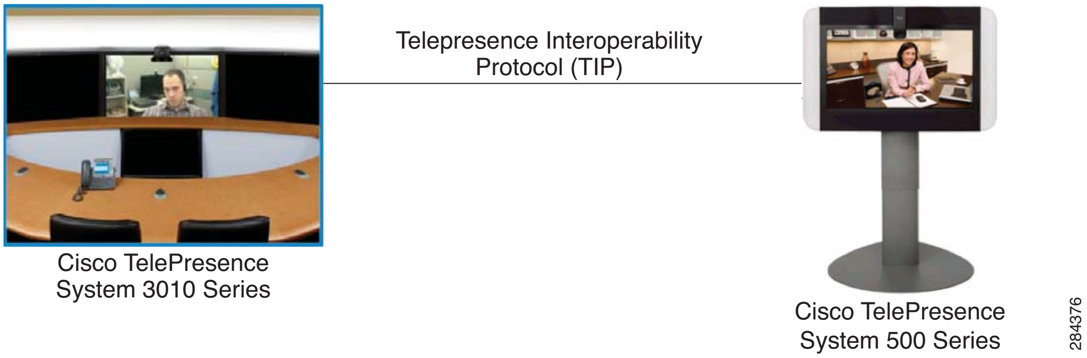
- Telepresence Interoperability Protocol (TIP), page 3-14
- ClearPath, page 3-15

Telepresence Interoperability Protocol (TIP)

Cisco originally developed the Telepresence Interoperability Protocol (TIP), but Cisco later transferred it to the International Multimedia Telecommunications Consortium (IMTC) as an open source protocol. The TIP standard defines how to multiplex multiple screens and audio streams into two Real-Time Transport Protocol (RTP) flows, one each for video and audio. It enables point-to-point and multipoint sessions as well as a mix of multi-screen and single-screen endpoints. The TIP specification also defines how Real-Time Transport Control Protocol (RTCP) application extensions are used to indicate profile capabilities and per-media flow options as a session is established. It also defines how devices can provide feedback and trigger resiliency mechanisms during the life of the streams.

As illustrated in Figure 3-11, TIP enables interoperability of multi-vendor, multi-screen IP video solutions by describing how switching of the screen (and its audio) should occur. TIP is used in video endpoints, video transcoders, video gateways, and MCUs (video conference bridges).

Figure 3-11 TIP Multiplexing in Action


Telepresence Interoperability Protocol (TIP)
Cisco TelePresence System 3010 Series
Cisco TelePresence System 500 Series
284376

ClearPath

Cisco ClearPath is a technology for removing the negative effects of up to 15% packet loss. It is a dynamic technology that combines a number of media resilience mechanisms. For example, when using lossy media, ClearPath helps to counterbalance the effects of the packet loss and thereby to improve the user experience. ClearPath is enabled by default and is used when it is supported on both ends of the video communication. The ClearPath mode is set by the **xConfiguration Conference PacketLossResilience Mode** command. All the media resilience mechanisms within ClearPath are H.264 standard-based, and the resulting encoded bit stream is H.264 compliant. ClearPath is designed to be independent of the call setup protocol, and it can be used by endpoints using H.323, SIP, and XMPP.

ClearPath uses the following technologies to produce the best possible user experience:

- Dynamic Bit Rate Adjustment, page 3-15
- Long-Term Reference Picture, page 3-15
- Video-Aware Forward Error Correction (FEC), page 3-15

Dynamic Bit Rate Adjustment

Dynamic bit rate adjustments adapt the call rate to the variable bandwidth available, downspeeding or upspeeding the call based on the packet loss condition. In the case of ClearPath, once the packet loss has decreased, upspeeding will occur. ClearPath uses a proactive sender approach by utilizing RTCP. In this case the sender is constantly reviewing the RTCP receiver reports and adjusting its bit rate accordingly.

Long-Term Reference Picture

Long-term reference frame recovery is a method for encoder-decoder resynchronization after a packet loss without the use of an I-frame. A repair P-frame can be used instead of a traditional I-frame when packet loss occurs, resulting in approximately 90% less data being transmitted to rebuild the frame.

A Long-Term Reference Picture (LTRP) is an I-frame that is stored in the encoder and decoder until they receive an explicit signal to do otherwise. For more information on Long-term reference frames or LTRPs, see the section on I-Frame, page 3-5.

Video-Aware Forward Error Correction (FEC)

Forward error correction (FEC) provides redundancy to the transmitted information by using a predetermined algorithm. The redundancy allows the receiver to detect and correct a limited number of errors occurring anywhere in the message, without the need to ask the sender for additional data. FEC gives the receiver an ability to correct errors without needing a reverse channel to request retransmission of data, but this advantage is at the cost of a fixed higher forward channel bandwidth. FEC protects the most important data (typically the repair P-frames) to make sure those frames are being received by the receiver. The endpoints do not use FEC on bandwidths lower than 768 kbps, and there must also be at least 1.5% of packet loss before FEC is introduced. ClearPath monitors the effectiveness of FEC, and if FEC is not efficient, ClearPath makes a decision not to do FEC.

CHAPTER 4

Call Control Protocols and IPv6 in IP Video Solutions

Revised: March 30, 2012, OL-27011-01

Protocols provide a complete set of specifications and suite of standards for communications between devices, This chapter does not discuss all the information available about the protocols but rather focuses on their most important features and characteristics in the context of handling video communications.

Call Control Protocols in IP Video Solutions

The primary call control protocols used in most IP video solutions today are H.323, Session Initiation Protocol (SIP), and Skinny Client Control Protocol (SCCP).

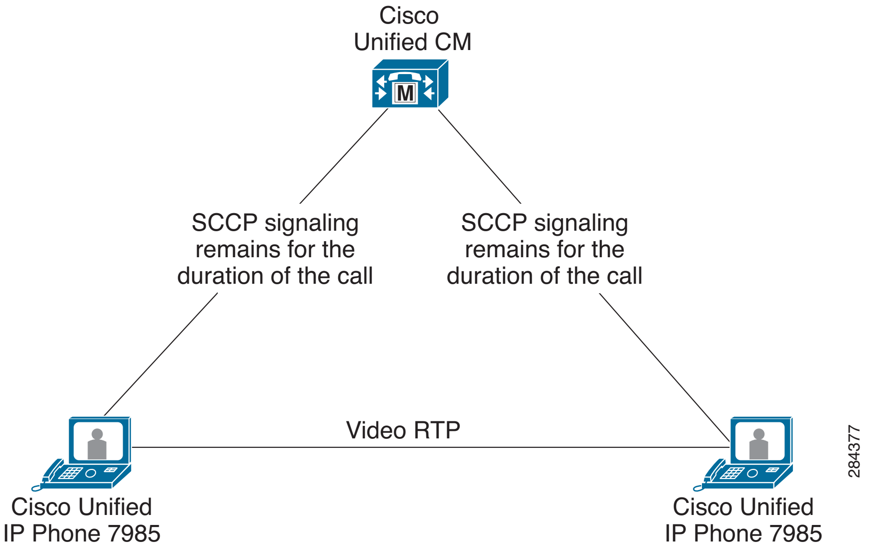
SCCP

Skinny Client Control Protocol (SCCP) was first developed by Cisco for IP Telephony applications. As IP Telephony matured, it integrated video as well and gave rise to Cisco IP Video Telephony. SCCP defines Transmission Control Protocol (TCP) as the transport protocol and a call agent in an architectural relationship with the endpoints (also known as a master/slave relationship). The call agent is the most fundamental difference between SCCP and the rest of the call control protocols discussed in this section. Because SCCP employs a central call agent, it inherently enables very advanced call functions for video endpoints that might not be available in other call control protocols.

Because SCCP defines a master/slave (or client/server) relationship between the call agent and the endpoints, the call agent must always remain available to the endpoint for call features to function. Therefore, SCCP might not be suitable for certain environments where the endpoints are expected to function independently from a call agent component.

Figure 4-1 illustrates the role of SCCP call control signaling in a deployment where Cisco Unified Communications Manager (Unified CM) is the call agent.

Figure 4-1 SCCP Signaling



As stated earlier, the SCCP specification provides support for advance call features in a video environment. Among those features, hold, resume, mute, and conferencing function exactly as they do for regular audio calls. The features that are most distinctive in SCCP endpoints are ad-hoc video conferencing and mute. Although support for ad-hoc video conferencing is not exclusive of SCCP, SCCP and the implementation of reservationless video conferencing in the video endpoints have made it easier for users to engage in ad-hoc video conferences. When the call control server is coupled with a compatible SCCP MCU, SCCP video phones are able to launch a conference by having users press a single key without making a previous conference reservation. This is an important difference from H.323, in which users must dial an always-on conference destination to establish a reservationless meeting.

The SCCP mute feature for video also functions differently than in other protocols. Unlike the mute function in H.323 and SIP, when mute is activated on the SCCP video terminal, both audio and video are muted simultaneously.

Just as SCCP enables intrinsically advanced call functionality on video endpoints through its phone-like technology and architecture, it also imposes some legacy phone-like behavior for video. Among the legacy behavior is the lack of support for uniform resource identifier (URI) dialing and data sharing. Therefore, when SCCP interoperates with some other protocol in a video deployment, it is important to consider any architectural limitations of SCCP. Table 4-1 lists other features not implemented in SCCP that should be considered when interoperating with H.323 or SIP for video.

Table 4-1 Features Not Implemented in SCCP

Feature not available in SCCP	Result	Workaround (if available)
Dynamic addition of video capability	Cannot promote an audio call to a video call	Ensure that video capabilities are available and broadcast at the beginning of the session
Far-end camera control (FECC) for SCCP endpoints	Cannot adjust remote cameras	Not available
Video codec renegotiation	Call session might be terminated if renegotiation occurs	Not available

The SCCP messages are encoded in hexadecimal, therefore reading them directly from the transmission is challenging. However, this encoding mechanism comes with the advantage that SCCP messages are generally smaller than with other call control protocols. For instance, an SCCP phone averages 256 bps in unencrypted call control traffic while a SIP phone averages 538 bps.

Another benefit of SCCP when used with video is that it allows authentication and encryption of media and signaling through Secure Real-time Transport Protocol (SRTP) and Transport Layer Security (TLS), respectively. When encryption is used, an SCCP video phone averages 415 bps while a SIP phone using encryption averages 619 bps.

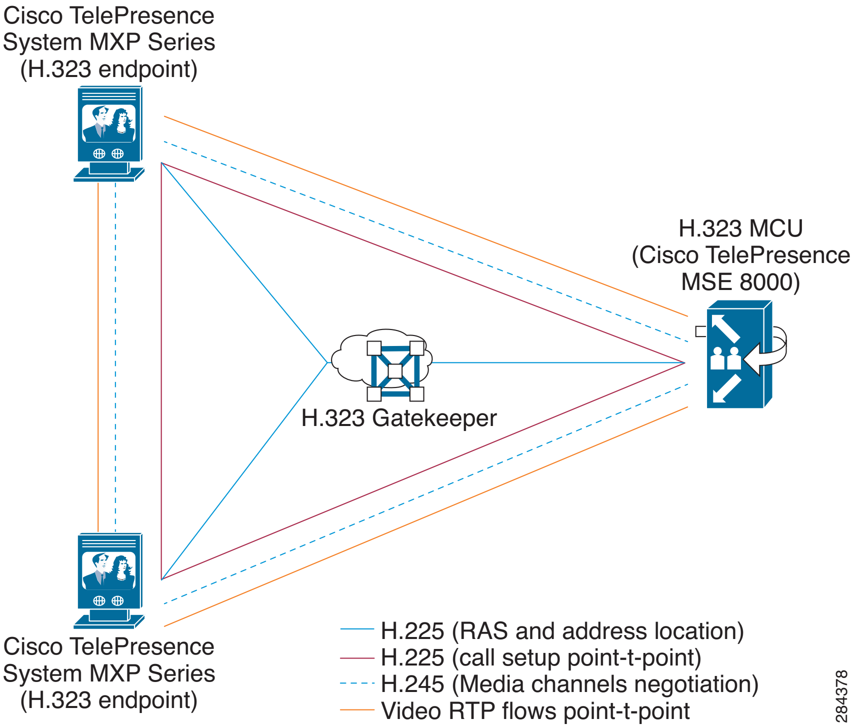
H.323

Unlike SCCP, H.323 is not a single standard or protocol but rather a suite of protocols and recommendations established by the International Telecommunication Union (ITU) Telecommunication Standardization Sector (ITU-T). H.323 is very strict in the definition of its features, expected behavior, and implementation, which puts H.323 in an advantageous position for interoperation between telecommunication vendors and providers. Because H.323 implementation is so well defined, it leaves very little room for misinterpretation of what is expected from the vendors when they interoperate.

H.323 uses a peer-to-peer protocol model that supports user-to-user communication without a centralized call control element. Because of H.323 robustness, it is not uncommon to find call control elements such as gatekeepers peered with endpoints from different vendors. As described earlier, H.323 is an umbrella protocol. An H.323 peer negotiates call setup and call admission control using H.225 and media channels using H.245. While H.225 uses User Datagram Protocol (UDP) and TCP as transport protocols, H.245 uses TCP only. Although this seems inconvenient for firewalls, H.323 is so well established in the telecommunications industry that most firewall vendors can efficiently inspect H.323 packets.

Figure 4-2 illustrates the use of H.323 with a gatekeeper as the call control element.

Figure 4-2 H.323 Signaling


Cisco TelePresence System MXP Series (H.323 endpoint)
H.323 MCU (Cisco TelePresence MSE 8000)
H.323 Gatekeeper
Cisco TelePresence System MXP Series (H.323 endpoint)
H.225 (RAS and address location)
H.225 (call setup point-t-point)
H.245 (Media channels negotiation)
Video RTP flows point-t-point
284378

H.323 provides strong support for a wide variety of video conferencing features, the most prominent of which are application sharing and far-end camera control (FECC). H.323 endpoints use H.224 and H.281 for FECC and H.239 for data sharing. FECC and application sharing in H.323 are key architectural differences between H.323 and other call control protocols. For instance, while SIP does not define how application sharing should be implemented, H.323 defines it clearly through Annex Q and its implementation of H.281 and H.224. With FECC in H.323, the camera control instructions are embedded into H.281 and later encapsulated in H.224, and RTP therefore provides a robust approach for transmission of the FECC instructions in the existing network infrastructure.

Application sharing is also very well defined in H.323, which uses H.239 to support it. H.239 defines how the management and addition of extra video channels must be implemented, and then the application video is sent over the additional video channel. Moreover, using a token system, H.239 ensures that only one participant at a time utilizes the application sharing functionality in the meeting.

Some features differ greatly between H.323 and other protocols. For example, some H.323 endpoints implement ad-hoc conferencing, but H.323 does not specify a central call control element in its architecture to execute conference resource tracking and establish conferences. Therefore, the ad-hoc conferencing behavior in most H.323 endpoints requires users to dial an always-on conference bridge.

Another example of protocol differences is that H.323 defines media encryption through H.235, but the definition of signaling encryption is not in the scope of H.323. Therefore, H.323 implementers commonly use either TLS or Internet Protocol Security (IPsec) when they need to secure the call signaling. This could potentially cause interoperability problems between endpoints from different vendors that use different approaches for securing the call signaling.

Although H.323 is highly evolved in its specification, H.323 features support only video and voice; they do not extend to instant messaging, presence, or other services. The lack of support in H.323 for new services should be thoughtfully considered when designing an IP Video network that might later integrate more communication methods besides voice and video alone.

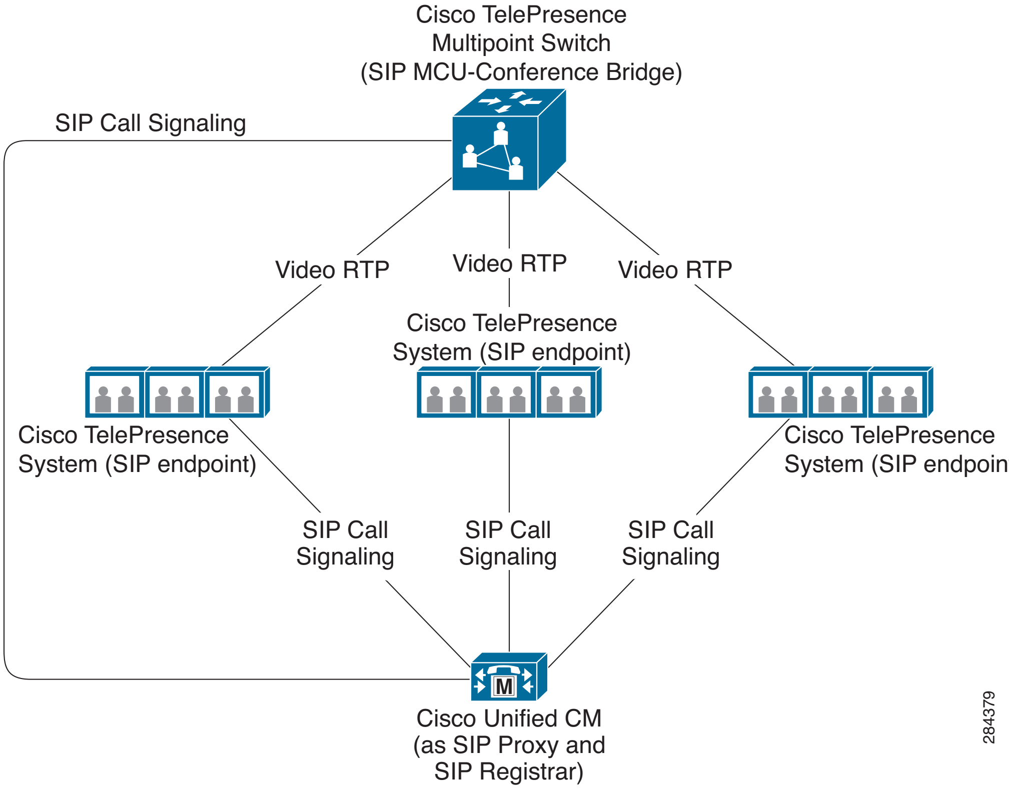
In addition, H.323 messages are encoded in binary, making it fairly challenging to interpret them without an appropriate dissector and potentially resulting in little-endian and big-endian errors when implementing the protocol messages. Although the H.323 messages are smaller than SIP messages, the difference in bandwidth can be considered negligible.

SIP

Session Initiation Protocol (SIP) is a peer-to-peer protocol. In the simplest implementation, SIP endpoints do not need a call control entity to contact each other, assuming they know their location. However, SIP also defines a client/server relationship so that the endpoints can make use of services, resources, and dialable destinations that are unknown to the endpoints. SIP is defined by the Internet Engineering Task Force (IETF) and is conglomeration of Requests for Comments (RFCs). Although the SIP core rules are defined by RFC 3261, the SIP standard includes more than a dozen RFCs.

In most enterprise deployments that use SIP, it is deployed with a call control element (client/server model) to provide a feature-rich experience, control over the dialable domains, and centralization of call control. SIP elements consist of two basic categories: user agent client (UAC) and user agent server (UAS). The element requesting connection to another element is the UAS, while the element receiving the request is the UAC. During a session, the same end-party can be a UAC for one transaction and a UAS for another, and the role is limited to a single transaction.

Figure 4-3 SIP Signaling


Cisco TelePresence Multipoint Switch (SIP MCU-Conference Bridge)
SIP Call Signaling
Video RTP
Video RTP
Video RTP
Cisco TelePresence System (SIP endpoint)
Cisco TelePresence System (SIP endpoint)
Cisco TelePresence System (SIP endpoint)
SIP Call Signaling
SIP Call Signaling
SIP Call Signaling
M
Cisco Unified CM (as SIP Proxy and SIP Registrar)
284379

In many regards, SIP is better categorized as a communications session signaling protocol than a telecommunications signaling protocol because SIP enables more than just the establishment of voice and video communications. SIP can enable instant messaging, presence, and so forth, whereas SCCP and H.323 are purely telecommunications protocols. Part of the strength of the SIP protocol specification to support a myriad of services comes from the fact that UAS and UAC elements must ignore what they do not understand or support. On occasion, however, this strength becomes one of SIP's disadvantages because it complicates interoperation between vendors. Furthermore, SIP is less detailed in its specification than SCCP or H.323, making vendor interoperation somewhat challenging at times. For example, in SIP there is more than one way to implement some features. If different vendors implement the same feature in different ways, they would be incompatible.

It is also important to note that some features defined in other call signaling protocols are either not defined in SIP or function differently than in the other protocols. For instance, before RFC 4353, there was no standard to define how ad-hoc conferencing should be implemented, and SIP implementers took different approaches to fill the void. In Cisco IP Video Telephony, ad-hoc conferencing was implemented by creating a proprietary approach using XML.

Another example of a gray area in SIP is application sharing. Some implementers use the 'm' (media-type) attribute to specify when application sharing media will be sent and when an additional video channel will be set up. However, SIP does not clearly define how these features should be implemented, which makes application sharing between SIP vendors challenging.

SIP is text-based and uses the ISO 10646 character set encoded in 8-bit Unicode Transformation Format (UTF-8). A SIP phone averages 538 bps for call control traffic in unencrypted mode, while an SCCP phone averages 256 bps. SIP can use either TCP or UDP. SIP implementations typically use port 5060 but SIP can also be implemented on a different port.

Call Control Protocol Selection in IP Video Solutions

Selecting the right protocol for the design of the IP video solution is crucial to its success. The wrong choice of protocol could result in scalability issues and/or the inability of users to execute expected features.

When selecting the call control protocol for an IP video solution or a call leg section, consider the following factors:

- What call features are currently needed by the users and what features are planned for the future? (For example, data sharing, encryption, and so forth)
- Which transport protocol (TCP or UDP) will be used? Some call control protocols are better suited to a particular transport protocol.
- Are network characteristics such as Network Address Translation Traversal (NAT-T) or deep inspection (security) needed? Sometimes video endpoints might need to be behind a firewall, NAT support might be required, or payload encryption might form part of the requirements.
- What interoperability will be needed (for example, with third-party H.323), and what type of endpoints and MCUs will be used? A particular call leg might include devices that do not support the protocol selected for the overall design. For example, interoperability might be required with the IP PBX where the audio deployment resides, and that IP PBX might use H.323.
- Will business-to-business (B2B) communications be needed? If so, will a B2B vendor be used or will there be a direct connection to a third-party company? If a B2B vendor is used, what call control protocol has the B2B vendor implemented?
- What are the application sharing requirements? For example, will H.239 be required?

The more information you can gather about the call control protocol usage and roadmap, the better the decision making process will be. Any additional information that is specific and relevant to you solution should be included in the criteria for protocol selection.

After gathering all the information needed for protocol selection, you can begin the selection process. When selecting the protocol, pay particular attention to the following areas:

- Scalability — Based on the information gathered, how much growth is expected in the IP Video solution deployment and how will that impact the protocol selected and the call control elements in the deployment.
- Use cases — Based on the call flows that are meaningful for the success of the deployment and the requirements gathered, real-world scenarios should be developed and careful studied to determine how the protocols affect them. For example, if users are expected to share applications through the video endpoints without access to laptops, that would reduce the protocol options to H.323 and SIP only.
- Customer requirements — Occasionally there might be requirements that do not fall into a clear use case or scalability area. Depending on the importance of the requirement, it can be assigned a certain weight for purposes of the protocol selection process.

IPv6 in IP Video Solutions

IP version 4 (IPv4) has by now exhausted all the public IP address assignments. There still is room in the private address ranges for a large enterprise to be able to increase its operations. Nevertheless, mobile devices are increasing exponentially the number of IP devices connecting in the enterprise, and as that trend continues, the implementation of IP version 6 (IPv6) eventually will be necessary to increase the number of available IP address assignments.

The risks of ignoring IPv6 and not planning ahead could range from inability to connect new devices to diminished business-to-business capabilities. Those risks, however, are in the mid-to-long-term future, given the fact that there are currently enough addresses for private usage.

Cisco already supports IPv6 in certain devices, such as the Cisco TelePresence C Series and Video Communication Server (VCS), while at the same time working on integration of IPv6 into the rest of its IP video portfolio. However, not many IP video equipment manufacturers support IPv6 at this time.

The best approach is to be prepared, know your network, and track the IP address assignments to understand when migration to IPv6 will be necessary for your network. When deploying a new IP video solution, ensure that the manufacturers and products you select have a clear roadmap for IPv6, and understand how much work the migration will take so that you can plan accordingly.

If your IP video solution requires internetworking of IPv4 and IPv6 devices, the Cisco VCS currently offers address translation between IPv4 and IPv6. For further information, refer to the documentation at

http://www.cisco.com/en/US/products/ps11337/prod_maintenance_guides_list.html

CHAPTER 5

Quality of Service and Call Admission Control

Revised: March 30, 2012, OL-27011-01

Data packets that make up the video streams are transmitted over the network. The packets reach their destination based on the order in which they are placed into the queue. In a simple network behavior, the first packet in is the first packet out. For a single LAN switch this process is fairly simple because the switch will have some ports that are sending the media and some that are receiving it. As the network grows, the scenario changes from being an ideal world of ordered packets to a mass of unordered packets, where there are more packets being generated at the same time than can be sent through network links that do not have sufficient capacity to handle the load. In these real-world scenarios, some methods are needed to control the flow of packets across the network links.

Quality of Service (QoS)

Quality of Service (QoS) is used to identify certain types of packets that can be processed ahead of others. The QoS information is inserted into the packets that need a different priority.

QoS can be compared to a freeway system and the vehicles that use it. The network is similar to the freeway system, which provides the vehicles (data packets in the case of QoS) with a way to travel from their starting point to their destination. As long as the freeway has sufficient lanes and there is no incident, traffic flows smoothly in most cases and the travel time is acceptable to most users. However, during peak traffic times, things might not be as good. Carpool lanes can help. Cars that meet certain criteria have the privilege to use the carpool lanes and bypass the traffic congestion. In addition, emergency services such as ambulances have an even higher priority to bypass other traffic. On the other hand, large or heavily loaded vehicles might use more lanes and can slow down traffic. QoS is similar in that it allows certain packets to have preferred access to the network and to be transmitted ahead of other packets in the queue.

Traditionally IP Precedence or Type of Service (ToS) (RFC 791) was specified using three bits in the IP packet. The Differentiated Services (DiffServ) (RFC 2474 and RFC 2475) model uses six bits and also maintains the IP Precedence values. The DiffServ model uses assured forwarding (RFC 2597) that defines various classes of traffic with a drop probability and expedited forwarding (RFC 2598) to provide for low loss, low latency, and low jitter service. The class in assured forwarding is used to group different types of traffic, and the drop probability is used to group the traffic that will experience dropped packets last. The expedited forwarding is used for traffic such as voice that is sensitive to packet drop and delays.

Each type of traffic can have a different QoS value, and the network then provides preference when it identifies packets that have a higher QoS value. Table 5-1 lists some of the standard values used for various types of voice and video traffic.

Table 5-1 Differentiated Services Code Point (DSCP) Values for Various Types of Traffic

Traffic Type	Layer 2 Class of Service	Layer 3 IP Precedence	Layer 3 DSCP
Call signaling	3	3	CS3 (24)
Voice	5	5	EF (46)
Video	4	4	AF41 (34)
TelePresence	4	4	CS4 (32)

Voice calls have only one stream of packets. Video calls have two streams, one for video and another for voice, and it is important for both streams of the call to have the same QoS marking.

Cisco Unified Communications Manager (Unified CM) supports endpoints that mark QoS for their media packets. Voice packets are marked as EF (DSCP value 46), while video devices mark media packets as AF41 (DSCP value 34) and TelePresence endpoints mark their traffic as CS4 (DSCP value 32). All call signaling is marked as CS3 (DSCP value 24).

QoS should be configured on the Cisco TelePresence Video Communication Server (VCS) because it processes the media and the call signaling. The endpoints that register to the VCS (such as the Cisco TelePresence System EX Series, C Series, Cisco IP Video Phone E20, or others) should be configured so that the call signaling uses DSCP CS3 and the media from those endpoints is marked as DSCP AF41.

Trust Boundary

Traffic on the network needs to be marked so that the network can trust it. The network elements such as the switches can be the trust boundary based on the switch that trusts the packets. It is important to establish a trust boundary so that the rest of the network does not have to remark packets for QoS. Access switches can trust IP phones based on association with them. Cisco switches use Cisco Discovery Protocol (CDP), a Layer 2 protocol, to associate phones with the switches. The switches use CDP to put IP phones in their respective voice VLANs. Access switches can then trust such phones for marking their packets with appropriate QoS and thus establish a trust boundary. When IP phones cannot be associated with the switch or when the trust boundary needs to be the access or distribution switches, the switch can build the trust boundary by enforcing the marking of packets based on criteria such as the IP addresses of the devices or the common ports used for signaling or media for calls.

Packet Queuing

While QoS helps the network distinguish different types of traffic and then prioritize it, the network uses a queuing mechanism to orderly move packets and control their flow. Queuing is widely used for networks that have low-capacity links between them, such as MAN or WAN networks.

Networks use the following common queuing mechanisms:

- First In, First Out (FIFO)

  This type of queuing is simple and gives the same preference to all packets based on the time they arrive in the queue. The packets are sent out through the switch or the network in the same order they arrive. This type of queuing is useful in networks that do not see a large change in the volume of packets.

- Priority Queuing (PQ)

  This type of queuing gives preference to packets with higher priority over packets with lower priority. The Priority Queue is commonly used for low-bandwidth traffic that is very sensitive to delay, such as voice calls.

- Weighted Fair Queuing (WFQ)

  This type of queuing uses multiple queues for priority and non-priority traffic. It provides for priority traffic without starving lower-priority traffic. This mechanism is used where networks have traffic that needs priority (such as voice traffic) as well as other important traffic such as business applications that should not be dropped.

- Class-Based Weighted Fair Queuing (CBWFQ)

  This type of queuing uses different classes to group traffic and then uses the WFQ mechanism while also providing dedicated bandwidth for some custom queues. This type of mechanism is used widely for interactive voice and video traffic when combined with business applications. This mechanism provides the advantage of flexibility for various types of enterprise deployments.

Queuing mechanisms provide Class of Service (CoS) for packets in the organization. The class of service is used to guarantee latency, jitter, and delay requirements. It also uses the link bandwidth more efficiently so that organizations can estimate the traffic they can send through their WAN links or plan their links to support desired traffic.

Policing traffic is important to prevent certain queues from using all the bandwidth. Policing prevents a certain type of traffic from exceeding its set use limit through a link. Traffic shaping and policing is needed to avoid packet drops and to allow servicing of non-critical traffic.

With video calls it is important for both the video and voice streams of the call to be sent through the same queue in order to avoid lip-sync issues. When the video and voice streams use different queues, one arrives later than the other and causes the video to be shown while the voice associated with that video may lag, or vice versa.

Call Admission Control

To ensure that the voice and video traffic does not use all the bandwidth in the link and cause other important data such as business applications to experience dropped packets, organizations can use calls admission control. Call admission control limits the number of calls allowed through a particular link between sites.

There are two main methods for limiting the number of calls on a link:

- Call counting — With this method, the call control agent counts the number of calls allowed between locations. Only calls of the same type are counted together, so a set voice codec and a set video codec make counting such calls easy.
- Bandwidth — This method is similar to call counting, but here the count consists of the amount of bandwidth used by the calls. The type of voice codec and the type of bandwidth for video calls are used to count the bandwidth.

Preserving the call quality is important. When calls traverse WAN links, oversubscribing the link can cause call quality to degrade. Call admission control is important because it can prevent calls from filling up the link. When routing calls, call control agents know if a call should be allowed or if the link cannot handle that call. This provides a consistent call experience to users.

Cisco Unified CM supports call admission control using the bandwidth method, so calls with different codec types and video bandwidth types can be supported on the enterprise network. Unified CM uses the mechanism of regions to specify per-call parameters for codec and video bandwidth. Unified CM also uses the mechanism of locations to specify the bandwidth value used to limit voice and video calls for a particular site. Cisco TelePresence VCS uses similar mechanisms, where Links set the per-call bandwidth and Pipes set the bandwidth for calls to a site.

Call counting and bandwidth methods for call accounting use static configured values, but the actual call may use less or more bandwidth. To find the actual bandwidth used for calls, Resource Reservation Protocol (RSVP) is used. This protocol looks at the actual path used by the media to determine if sufficient bandwidth is available for a call. When a device in the path does not have the bandwidth to service the call, that condition gets reported through RSVP and the call might not be allowed.

Cisco Unified CM uses a separate media device called Cisco RSVP Agent to negotiate the network bandwidth, using RSVP on behalf of the video endpoints. This allows for a more accurate accounting of calls through it.

TelePresence calls do not use call admission control because the network is designed to allow the needed traffic so as to guarantee the user experience through such technologies. TelePresence traffic is marked differently than voice and video; thus, it can be queued separately and can provide low latency and delay, thereby preserving the user experience.

When there is more than one call control agent in the organization, each agent does its own call admission control. They work in parallel and do not know about calls through each other. If a call occurs between two different call control agents at the same site, both agents may count the bandwidth for that call even though it does not use any WAN bandwidth. Therefore, it is important to have only one call control agent doing call admission control for the devices in the organization.

CHAPTER 6

Dial Plan

Revised: March 30, 2012, OL-27011-01

Land-line and cell phone users dial a number to reach each other. Number-based dialing is commonly known as E.164 or PSTN dialing. PSTN numbers have various components such as country codes, area codes, and the destination. PSTN service providers then resolve the dialed numbers as local, long distance, or inter-country numbers and connect the two parties. Similarly, organizations use prefixes such as 9 or 0 for an outside dial tone to differentiate services. Users dial this prefix and then the number they wish to call. Dialing internal numbers within an organization may also use prefixes so that users can dial short numbers and not the entire E.164 number. Some organizations use an abbreviated dial plan, where users in the organization dial a 4 or 5 digit number to reach their colleagues within the same company.

This forms the basis of the dial plan that connects various enterprises, mobile networks, and the PSTN so that people on the respective networks can reach each other. Organizations that are conscious of costs of long distance or international calls can use a numeric dial plan to restrict calls and allow only certain users to dial long distance or international. Similarly, organizations need to have the ability to block unwanted calls such as calls from telemarketing agencies. Organizations should also consider ways to prevent toll fraud when using a number-based dial plan.

Modern Internet communication has been popular since the introduction of e-mail. E-mail addresses provide a way for users to be identified uniquely and are now the new identity. This identity can be used to reach users when they are represented as Uniform Resource Identifiers (URIs). Instant messaging for people to chat with each other uses URIs, and Voice over IP (VoIP) can also use URIs to send calls. The format for a URI is <*user-id*>@<*domain-name*>. This format is simple for users and generally easier to remember than a number.

An advanced system is needed to locate the entity represented by the URI, so that calls can be routed to it or the device using it. The system must be able to identify the domain and the servers that represent that domain, as well as being able to service calls. At a minimum, external dependencies such as Domain Name System (DNS) need to be considered. Various VoIP protocols such as H.323 can use Electronic Numbering (ENUM) for this purpose, but this method is not widely used. SIP has made URI dialing more popular due to its origins in web technologies. While different types of dialing methods offer users and enterprises a flexible way to route calls and provide reachability, they require some method to provide good translations between the systems that provide these different types of dialing experiences for users.

Dial Plan Dependencies

When discussing the wide reach of the PSTN, it is important to understand the dependencies of the dial plan:

- E.164-based dial plans can use a variable number of digits. Service providers and telephony systems should be able to accommodate variable-length dialed numbers, and users should be used to dialing those numbers. If there are different lengths of dialed numbers, the way the service provider handles them will be different than how an organization would handle them. Service providers need to normalize the numbers for caller ID, while the organizations have to normalize the numbers for direct inward dialing (DID) or for a country-specific dial plan and thus might have to transform the calling or called numbers.
- Connecting a voice and video system to the rest of the world requires PSTN or IP connectivity to call other devices and to be reachable. IP-based systems need connectivity to the external networks and may use direct peering or some other mechanism to resolve the dial plan through the IP network.
- IP-based networks need different mechanisms to resolve dial plan between them.
  - Electronic Numbering (ENUM) is used for number-based calling. ENUM provides a way to translate an E.164 number to a corresponding IP address. The service is provided by a ENUM server. Subscription to the service and network connectivity are needed, and DNS services are also needed.
  - DNS Service (SRV) is used for systems supporting URI dialing. The DNS SRV records provide domain name information as well as information on call services such as H.323 or SIP protocols and UDP or TCP types and additional parameters that are useful for call agents. Call agents then use this information to send calls to hosts.
- A single user may have more than one endpoint or device. In such cases the dial plan should be able to reach all of the user's devices when a single number or URI is dialed. Therefore, number-based systems should support shared lines and URI-based systems should support alias.

Number-Based Dial Plan Networks

Numeric dial plan systems are more widely deployed. PSTN and H.323 systems use numbers to reach endpoints. PSTN and IP PBX systems own their dial plan and assign numbers to their endpoints. The H.323 network gets the dial plan from the endpoints because the endpoints own the dial plan and share it with the system so that calls can be routed. Exact or most significant match of dialed digits is used to resolve the dial plan for such systems.

The emergency services network such as the E.911 network in the United States is another type of system that uses numeric dialing. Numeric dialing is used not only to reach the services but also to reach the callers, so that additional information associated with the location of the caller can be used to provide the emergency services.

URI-Based Dial Plan Networks

Software-based video endpoints make it easy to dial URIs. Most applications make it even easier by providing the click-to-call feature. The ease of dialing based on e-mail IDs is very appealing to users. Most deployments use a common directory service such as Lightweight Directory Access Protocol (LDAP) to reference a common identity system that provides e-mail IDs for users to dial each other. Therefore calling outside of the organization's LDAP system has external dependencies.

DNS provides information on domains so that calls can be routed using domains, similar to instant messaging systems. DNS is the most common system that is widely used for business-to-business communications as a common network where information on how to connect to the organizations is available, so that organizations can make external calls to each other.

Call Resolution with Dial Plans

Call agents use the dialed information to make decisions about how to route calls. Common ways of processing numbers for calls use the best match for the digits or the most significant match of the routes that the call agent handles. Endpoints have unique E.164 numbers. Call agents can identify the country-specific information or the prefix information and then make decisions to send calls to respective endpoints or to neighbors who service those dialed numbers. Call agents also need to resolve calls for emergency services and for local calls, so that calls can then be sent to local gateways. In the case of numeric dialing, the concept of local area codes serves as a way to identify calls that need to be sent to local resources or to systems that are not local.

Registering devices with URIs also provides uniqueness for users. Call agents can then identify the destination based on the URI. Multiple servers across multiple locations can service a domain, and therefore the call agents must be able to resolve the domain and then the server that has the registration for a given URI. If a destination location is not available, other methods will be needed for making decisions to resolve calls for emergency services, local IP trunks, or hop-off gateways. In addition, call agents that need to route with URIs for conferencing services should be capable of choosing the conference bridge closest to the participants.

PSTN Access

Communicating to the rest of the world means that there should be an interface to the PSTN and the IP network. Gateways that connect to the PSTN through ISDN provide for voice connectivity to the PSTN. The gateways provide key functions of normalizing calling and called numbers when interfacing with a numeric-based PSTN network. Toll fraud prevention should be used for PSTN gateways.

Border elements or IP gateways provide connectivity to an IP-based PSTN either through the service provider or through the public Internet. The gateways provide topology hiding for the internal IP networks and normalization for calls. The gateways need to have additional security measures to prevent denial-of-service attacks and attempts to compromise them.

Transformations

Call agents perform transformations on the dial plan by transparently changing the calling and called numbers to normalize information of the callers. Transformation is a key function of call agents to standardize user information for calls. It is also important for relaying information for calls that are not answered, so that users can then dial back to the callers once they have time to make the calls. Organizations use transformations to mask their caller IDs or to present callback information to a generic and standardized format. For PSTN calls, organizations use their public numbers for outgoing calls rather than expose the internal numbers of their users. URI-based systems use transformations to normalize calls within their domain and for calls from external organizations.

Some organizations also use DID to oversubscribe internal users to their external ISDN lines. This gives the organization a way to provide endpoints to all its users while choosing appropriate lines for external calls. With URI registration, the concept of oversubscription is limited to the capacity used by calls rather than the number range of the endpoints.

Dial Plan Manipulation

Changing various aspects of the caller or called-party information at various system elements is done to assist call routing. Caller information needs to be manipulated so that information presented to the caller or called party is normalized. It also is useful if the call goes unanswered and the user needs to reference information on missed calls or received calls. In the case of numeric dial plans, these goals are simple to address; however, for URI dialing the call agent must have mechanisms to achieve the desired goals.

Call agents may manipulate caller information to perform call resolution. Call agents can choose to add area codes to change numbers to E.164 format and then route them. Providing such manipulation may require basic pattern matching and replacement functionality. In the case of URI dialing, normalizing the domain may be used so that call agents can process the URI. To provide searching and replacement of alphanumeric content, a more elaborate method using regular expressions is needed. Regular expressions provide greater flexibility for alphanumeric URI formats, but they might not be as simple as numeric dial plan manipulation.

Classes of Restrictions

Organizations typically want to allowing some users to have access to certain services while denying other users access to those same services. The most common cases are for long distance and international dialing restrictions. While restrictions on calls and callers are easy to implement and deploy based on caller number or called number, those restrictions can be complex for URI dialing. Restrictions can also be used to prevent calls to toll-free destinations or from telemarketing agencies. Numeric dial plans can use know number ranges, such as 1-888 collect call numbers, to block calls. With URI dialing, this can be a challenge because restricting calls based on URIs requires a list of restricted URIs that need to be configured through the call agent to define a class-of-restriction policy for calls.

CHAPTER 7

Deployment Guidelines for Video Networks

Revised: March 30, 2012, OL-27011-01

When deploying a video network, it is imperative to design, plan, and implement the video network in a way that provides the best user experience possible. Video is an application subject to the perceptions of the various parties during a call. If one of the video users in a meeting or collaboration effort does not have a satisfying experience, that perception can easily be transferred onto the rest of the attendees.

The following sections offer general guidelines for designing video networks:

- Planning a Deployment Topology for Video, page 7-1
- Allocation of Video Resources, page 7-10
- Creating a Video-Ready Network, page 7-14
- Integration with Standalone Video Networks, page 7-19

Planning a Deployment Topology for Video

When deploying video applications, it is fundamental to identify and plan the topology or topologies that are deployed or that should be deployed to cover the needs of the organization. Current video applications use the following main video deployment topology models:

- Intra-Campus, page 7-2
- Intra-Enterprise, page 7-3
- Inter-Enterprise (Business-to-Business), page 7-4

Video applications also use the following main call processing models:

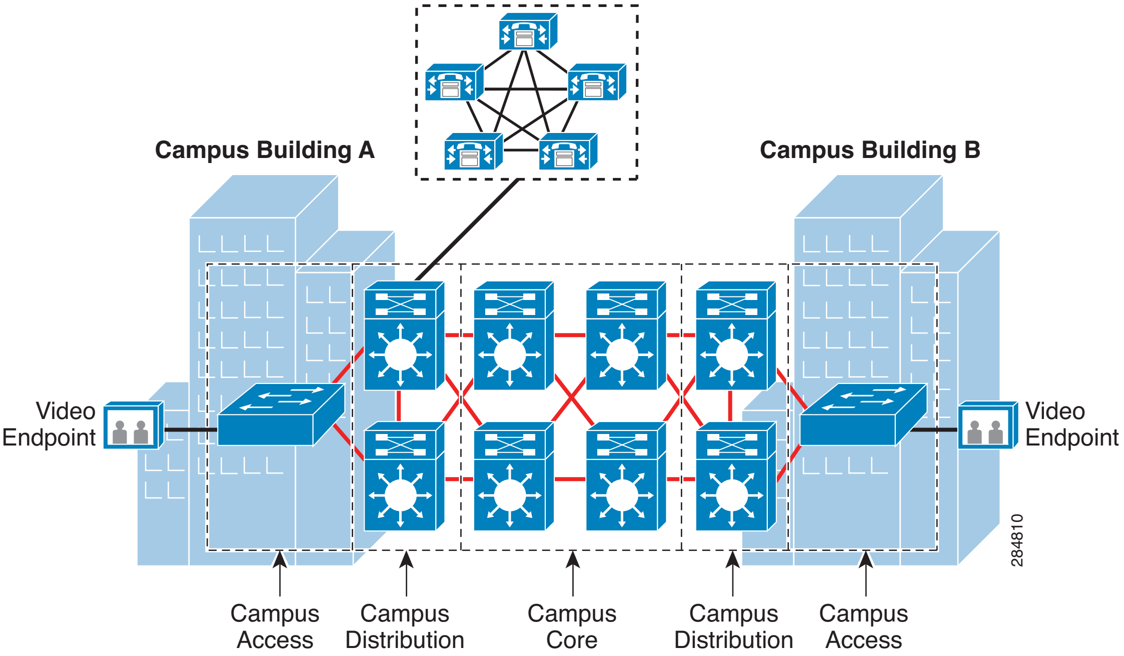
- Single-Site Call Processing, page 7-5
- Multi-Site Call Processing, page 7-5
- Hosted Call Processing (Video as a Service), page 7-7

The following sections provide an overview of these deployment models as well as guidelines for choosing a call processing model and topology.

Intra-Campus

An intra-campus topology for video is limited to offering video throughout a single company site or campus. This topology model is better suited for companies that require increased meeting effectiveness and productivity without requiring the users to move throughout the facilities. The intra-campus video deployment model can be used in conjunction with the intra-enterprise and inter-enterprise topology models to meet the needs of the organization. Figure 7-1 depicts an intra-campus topology using single-site call processing.

Figure 7-1 Intra-Campus Deployment with Two Buildings (Everything in One Building Except Video Endpoints)



With respect to call processing, the single-site and hosted call processing models are a better fit for intra-campus video topology models. Deciding between the two depends greatly on the endpoint density, planned growth, features required, and cost.

For instance, hosted video call processing deployment models offer a very feature-rich experience with a low cost investment. However, certain local multipoint call flows might require the media streams to travel outside of the premises when embedded video resources are not available or when their capacity is exceeded, and therefore the bandwidth-to-cost relationship becomes a crucial factor as user density increases.

On the other hand, a single-site call processing model has a higher initial investment because features require either hardware or software licensing for them to be available. Nevertheless, the single-site call processing model allows for user growth with a lower cost ratio after it has being deployed.

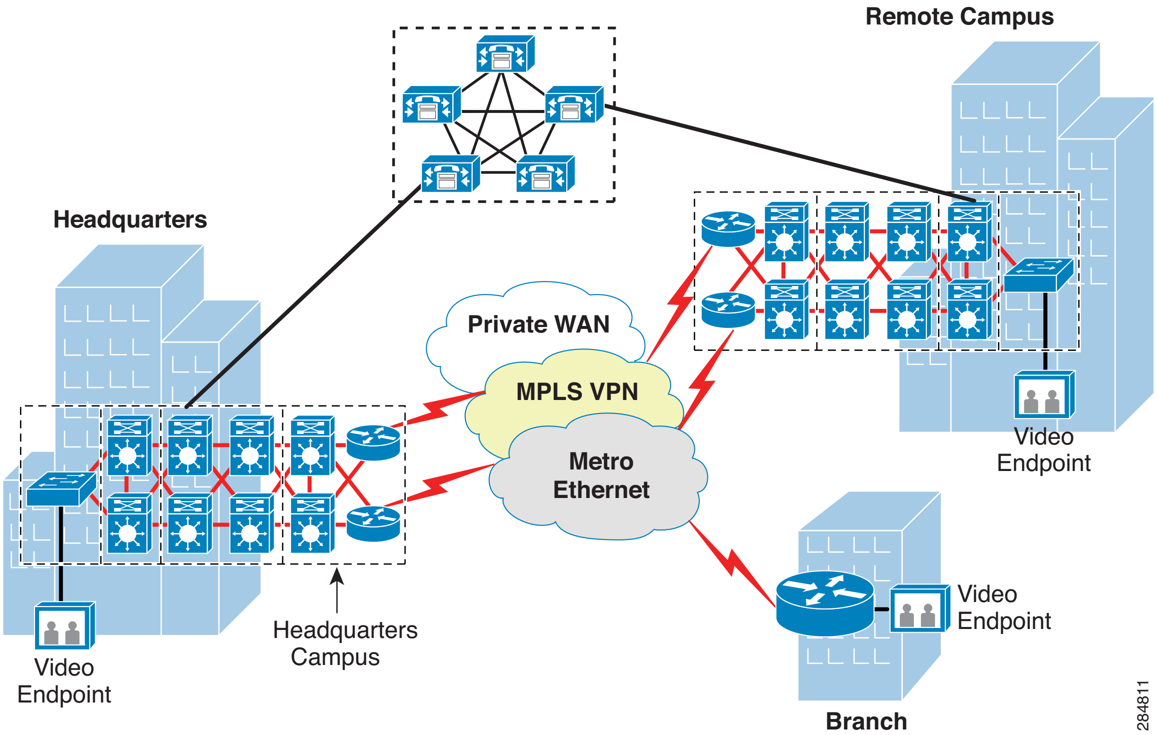
Intra-Enterprise

An intra-enterprise topology for video enables more than one site within the same company to use video applications through the WAN. The intra-enterprise deployment model is suitable for businesses that often require employees to travel extensively for internal meetings. Deploying video within the enterprise not only improves productivity by saving travel time and providing feature-rich collaboration, but it also reduces travel expenses. Furthermore, the overall quality of work and life is often improved when employees have to travel less.

To enable video across multiple sites, the intra-enterprise topology must make use of high-speed supporting WAN links to provide a rich video experience. For details about dimensioning the WAN links, see Scalability and Performance, page 7-16.

Figure 7-2 illustrates an intra-enterprise topology that uses a multi-site call processing deployment model.

Figure 7-2 Multiple Sites with Distributed Multi-Site Cisco Unified Communications Managers



Intra-enterprise deployments can use either multi-site or hosted call processing deployment models. As with intra-campus deployments, the same considerations apply when deciding between the two call processing models.

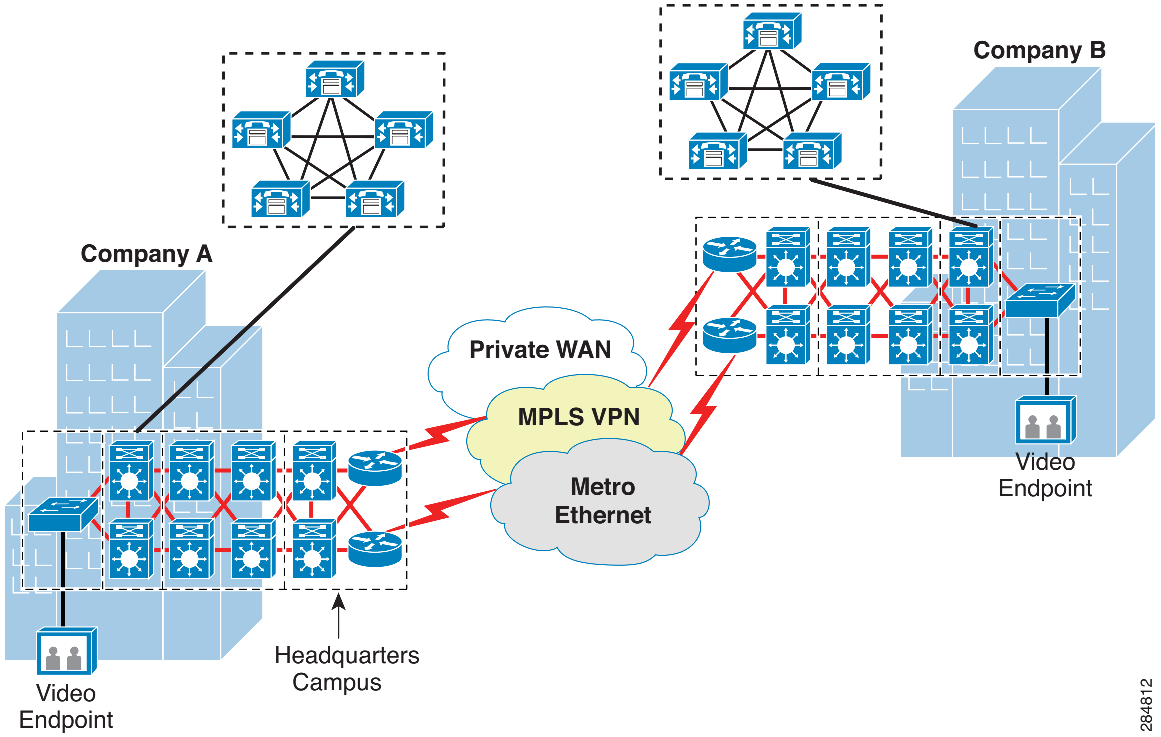
Inter-Enterprise (Business-to-Business)

The inter-enterprise network deployment model not only connects video endpoints within an enterprise, but also allows for video endpoints within one enterprise to call systems within another enterprise. The inter-enterprise model expands on the intra-campus and intra-enterprise models to include connectivity between different enterprises. This is also referred to as the business-to-business (B2B) video deployment model.

The inter-enterprise model offers the most flexibility and is suitable for businesses that often require employees to travel extensively for both internal and external meetings. In addition to the business advantages of the intra-enterprise model, the B2B topology deployment model lets employees maintain high-quality customer relations without the associated costs of travel time and expense.

Figure 7-3 depicts two companies communicating with each other through an inter-enterprise topology and single-site call processing.

Figure 7-3 Inter-Enterprise (B2B) Deployment with Single-Site Call Processing at Each Company

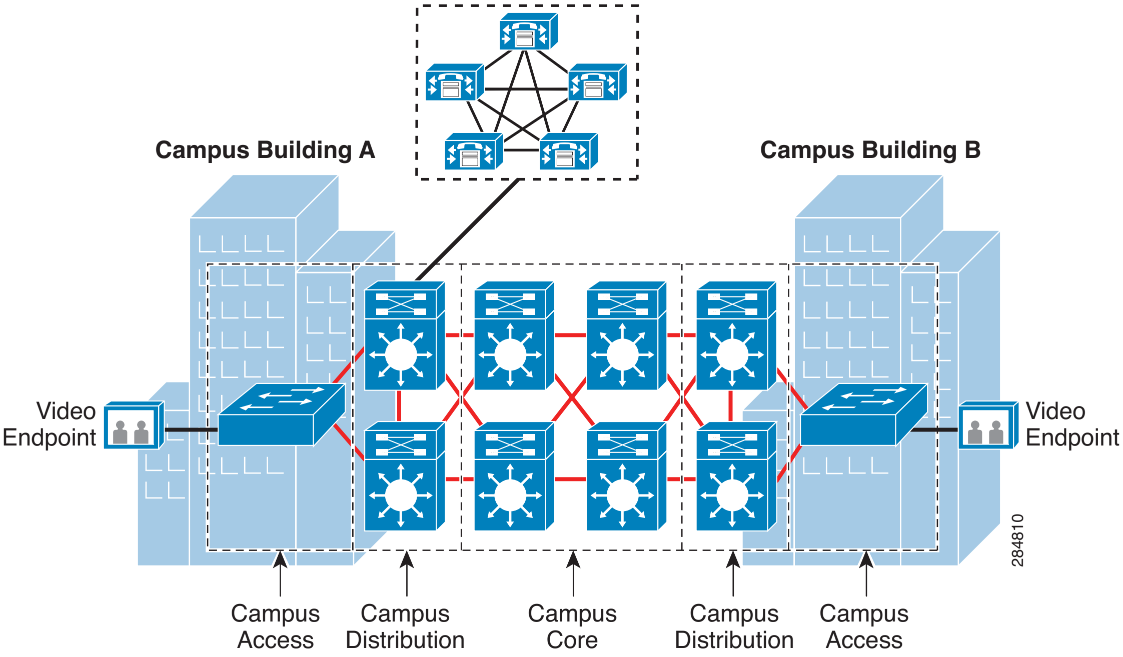


All three call processing models can work with the inter-enterprise deployment model, and the same considerations apply as with intra-campus and intra-enterprise deployments.

Single-Site Call Processing

A single-site call processing model confines call processing to service a single site, and the call processing agents are in the same location as the serviced endpoints. Whatever the distance is between the call processing agent and the endpoints, it should be serviced by LAN speed links. The single-site model is suitable for medium-sized businesses and government operations that reside at one site and that have basic video call processing needs, but where growth might be explosive or where the user density is very high, thus making the bandwidth-to-cost ratio of a hosted solution prohibitively expensive. Figure 7-4 depicts an intra-campus topology using single-site call processing.

Figure 7-4 Intra-Campus Deployment with Two Buildings (Everything in One Building Except Video Endpoints)

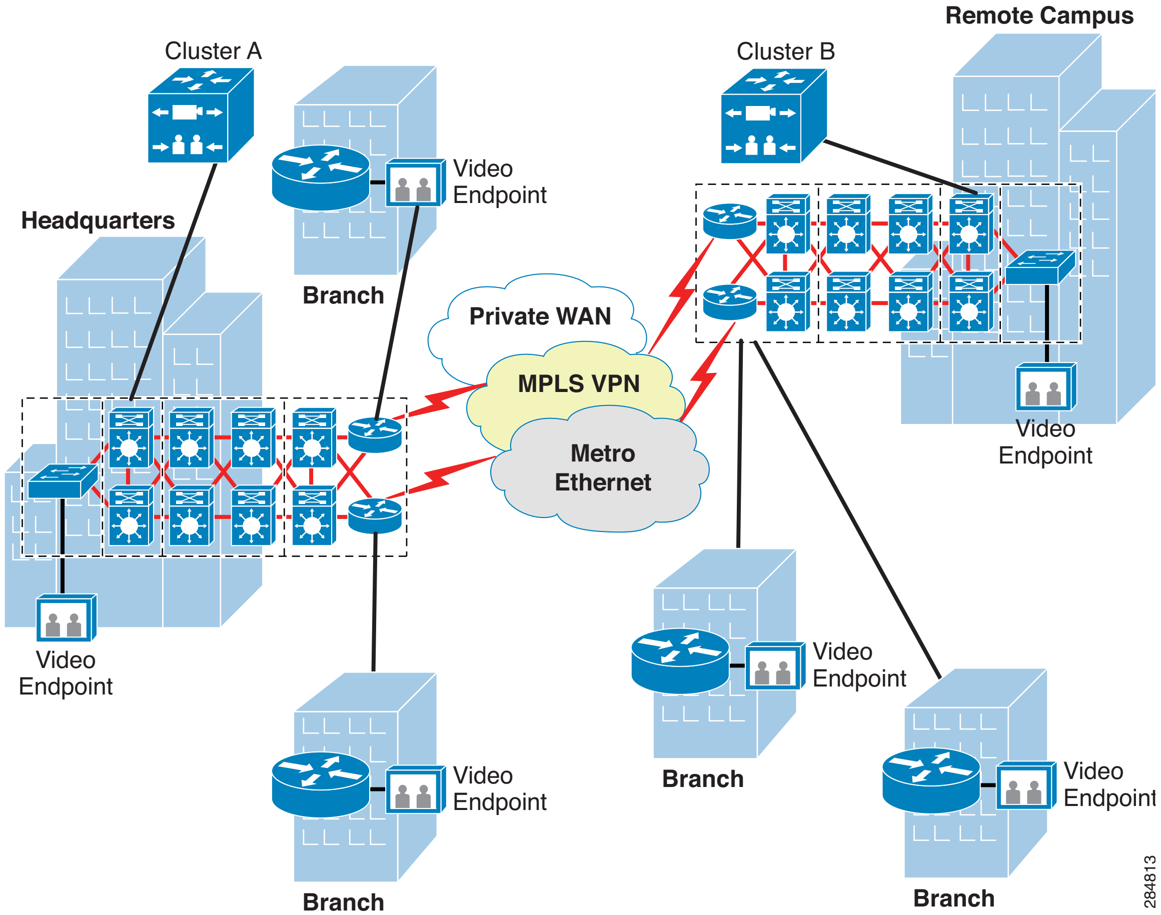


Multi-Site Call Processing

In a multi-site call processing model, the call processing agents can all be at the same location (centralized multi-site call processing) or distributed across various locations where high video user density exists or where service is critical and a backup is required.

Within the same call processing agent cluster, multi-site call processing can service a variety of topologies (for example, hub-and-spoke and multiple-hubs-to-spokes topologies) using either a centralized or distributed multi-site call processing model. Figure 7-5 illustrates a distributed call processing model where a multi-site call processing deployment services a large central site and multiple remote or branch sites, with the home office sites or smaller branches depending upon the larger ones (multiple hubs to spokes) for call processing services.

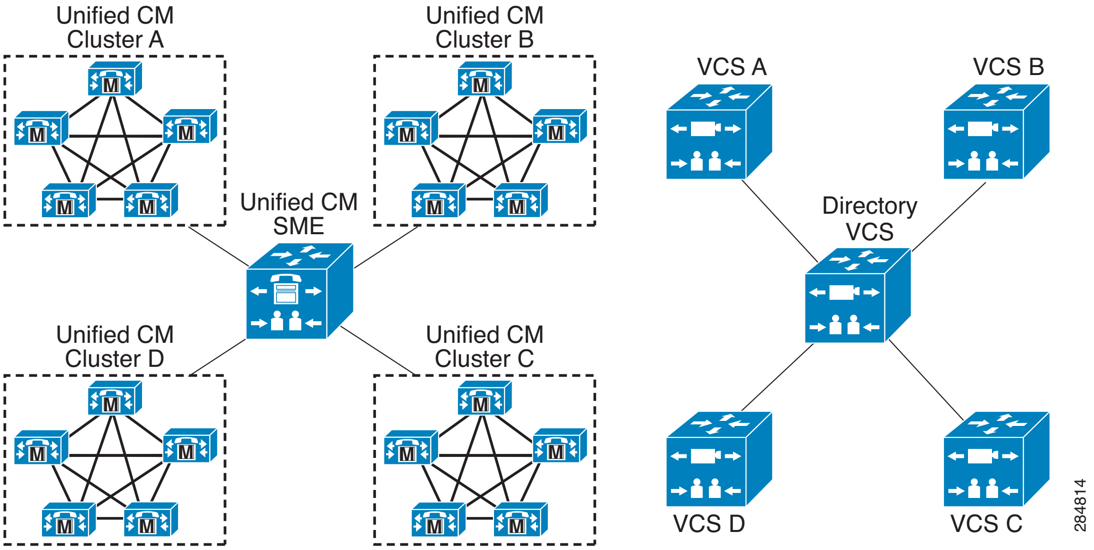
Figure 7-5 Distributed Multi-Site Call Processing



The multi-site call processing model also includes deployments where clustered call processing agents interact with other clustered call processing agents through a call processing element deployed for the sole purpose of aggregating the call routing between the clusters. Deploying an aggregating call processing entity is advantageous because it eliminates the need to implement full-mesh connectivity between all the call processing clusters. Instead, the various leaf clusters engage with the aggregating call processing element when the leaf clusters communicate with each other. If the dial plan is implemented correctly to provide hierarchy and allow for increasing the capacity of the overall solution, no dial plan updates are needed in the leaf clusters when new leaf clusters are added.

Figure 7-6 illustrates two examples of multi-site aggregated call processing deployments, one using Cisco Unified Communications Manager Session Management Edition (SME) the other one using a Cisco TelePresence Video Communication Server (VCS) as a directory gatekeeper.

Figure 7-6 Two Examples of Multi-Site Aggregated Call Processing Deployments



Note Although H.323 gatekeepers do allow the concept of a cluster (alternate gatekeeper), a single gatekeeper is also considered as a clustered call processing agent for purposes of the above discussion because H.323 gatekeepers are self-deterministic call routing entities.

Hosted Call Processing (Video as a Service)

Hosted deployment models refer to services provided and managed from the cloud. The offering is compelling because of the lower cost of ownership (low investment) and feature-rich experience it provides. Hosted video solutions are aimed at mid-sized and smaller businesses, providing them with an affordable entry point into the world of enterprise-grade video, and some hosted solutions also provide a migration path that protects the original investment as the business grows.

In this deployment model, more than just the call processing elements are often off-premises, thus requiring that the video streams travel off premises in some multipoint call flow scenarios. In these cases, higher user density in a service location will require higher bandwidth to service them.

Guidelines for Choose a Call Processing Topology and Video Endpoints

Choosing the right elements and deployment models to implement or expand a video network is instrumental for ensuring that the desired features, performance, and scalability are achieved. Moreover, choosing the wrong elements and models for the video network can result in costly changes to provide the functionality that the organization requires.

The following sections provide general guidelines for selecting the elements and deployment models for a video network:

- Call Processing Model and Call Processing Agent Selection Guidelines, page 7-8
- Endpoint Selection Guidelines, page 7-9
- Design Considerations for Video Networks, page 7-9

For information about call signaling protocol selection, see the chapter on Call Control Protocols and IPv6 in IP Video Solutions, page 4-1.

Call Processing Model and Call Processing Agent Selection Guidelines

To choose the correct call processing agent and its deployment model and topology, it is important to consider the following points in addition to the requirements of the organization during the design phase of the deployment:

- What features are needed to fulfill the use cases for the success criteria of the deployment? For example, SIP, SRTP, BFCP, or IPv6 video support.
- Are any video endpoints already deployed in the network that will have to be serviced? Are there requirements for previous video endpoints that have to be serviced by the newly selected call agent, interoperation with a previous video network, and so forth?
- If there are any previous video elements to be included in the deployment, what protocols do they use or can they use? For example, multi-protocol endpoints (SIP and H.323) or single-protocol endpoints.
- Has the call control protocol been selected? If not, are there any features dependent on a particular call control protocol? For example, H.239 can be used only in conjunction with H.323.
- What are the locations that will need video service? What is their user density? How critical will redundancy be for each particular location? It is advisable to provide a call agent for redundancy purposes at each site that has a high user density (distributed call processing). A very high user density will also create challenges (internet bandwidth usage) under certain call flows for hosted call processing.
- Will protocol interworking be used? Interworking will significantly influence the placement of the call agents under certain circumstances because the media stream might have to traverse the call processing agent in order to reach its destination.
- What is the maximum number of video calls you expect to service? A single Cisco TelePresence Video Communication Server (VCS) 7.0 has a limit of 500 non-traversal calls. Beyond that, a second VCS is needed (in a cluster or standalone) to service the rest of the calls.

The outcome of these considerations, coupled with the customer requirements, product data sheets, and product release notes, will determine which call processing agent to select and which call processing model and topology to use. For example, if the requirements include the use of BFCP as an application sharing technology and service for a maximum of 2000 calls, then a VCS cluster would be the best choice.

Endpoint Selection Guidelines

Selecting the right endpoints for the job is just as important as selecting the call processing agent. In addition to the customer requirements, the following points help determine the selection:

- What call control protocol will be used, H.323 or SIP?
- Will embedded video resources be needed for video conferencing? If so, Cisco TelePresence System EX90 would be a suitable choice.
- What video resolution formats will be required? For example, HD 720p.
- What other endpoints will be engaged in a call with the given endpoint(s) being selected? For example, Cisco Unified IP Phone 9971.
- What application sharing technologies will be needed? For example, BFCP over UDP.
- What are the mobility requirements? Will this endpoint be a mobile endpoint (collaboration tablet)?

The outcome of these considerations, coupled with the customer requirements, product data sheets, and product release notes, will determine which endpoints to select.

Design Considerations for Video Networks

When designing a video network, it is important to consider the implications of interworking. Depending on the call processing agent selected, the media streams might have to traverse the call agent when interworking is used. Therefore, interworking might have a negative impact on bandwidth usage if the call processing agent is remote to the endpoints engaged in the call because the call would not flow point to point.

Additionally, although DNS SRV records are typically for scalability purposes (the number of SIP trunks required to integrate a system is reduced when using SRV records), call processing agents behave differently with regard to endpoint registration and call processing peering when it comes to the use of DNS SRV records. These differences can create unexpected conditions when integrating different call agents if the behavioral differences are not understood prior to the integration.

Allocation of Video Resources

Whether servicing one or many physical locations, there are a number of considerations that need to be weighed to determine the best network locations for the video resources. Video resources can be either dedicated or embedded. Embedded video resources lie inside of the endpoints and service calls only for the endpoints that contain those resources. Dedicated video resources, on the other hand, reside in appliances separate from the endpoints, and they service any endpoints that have access to those resources.

Correct distribution of the video resources is necessary to achieve the desired level of user experience and, more often than not, the right level of redundancy and availability. The more factors you take into consideration, the more reliable your determination will be for the locations of the video resources. The following factors should be integrated into the decision making process to determine the best allocation model and locations for the video resources:

- Branch available bandwidth
- Bandwidth cost
- Remote video resource cost
- Usage patterns at headquarters and remote sites
- Call agent bridge selection algorithm
- Type of video resources

The following basic models can be used to allocate dedicated video resources in a deployment:

- Centralized Video Resource Allocation, page 7-10
- Distributed Video Resource Allocation, page 7-13

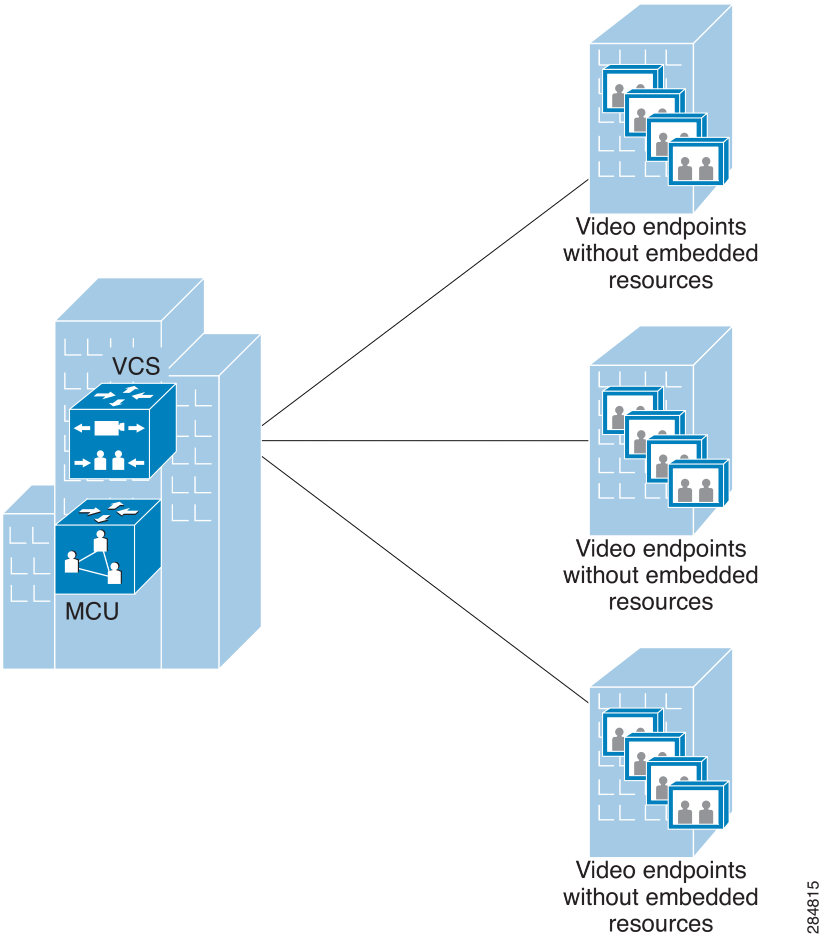
You can also combine embedded video resources with these models for dedicated video resources to form a hybrid model, if needed, to fit the necessities of your video solution more precisely.

Centralized Video Resource Allocation

Centralized resource allocation should be considered when the combined costs of placing the resources in the same location are less than distributing them. The feasibility of a centralized resource allocation should also be considered. For example, not all scenarios will be suitable for a centralized resource architecture if the concentration of the resources induces an undesirable condition on the endpoints (for example, unacceptable jitter).

As previously stated, the factors listed in the section on Guidelines for Choose a Call Processing Topology and Video Endpoints, page 7-8, should be integrated into the decision making process to select the best approach to fit a given scenario. For instance, consider the scenario depicted in Figure 7-7. Although at first it might seem less costly to concentrate all the video resources in the headquarters, such resource concentration would create the side effect of increasing the bandwidth requirements for the WAN links between the hub and spokes. Furthermore, multipoint conference capacity would be limited in the remote sites by the bandwidth provided in the WAN links.

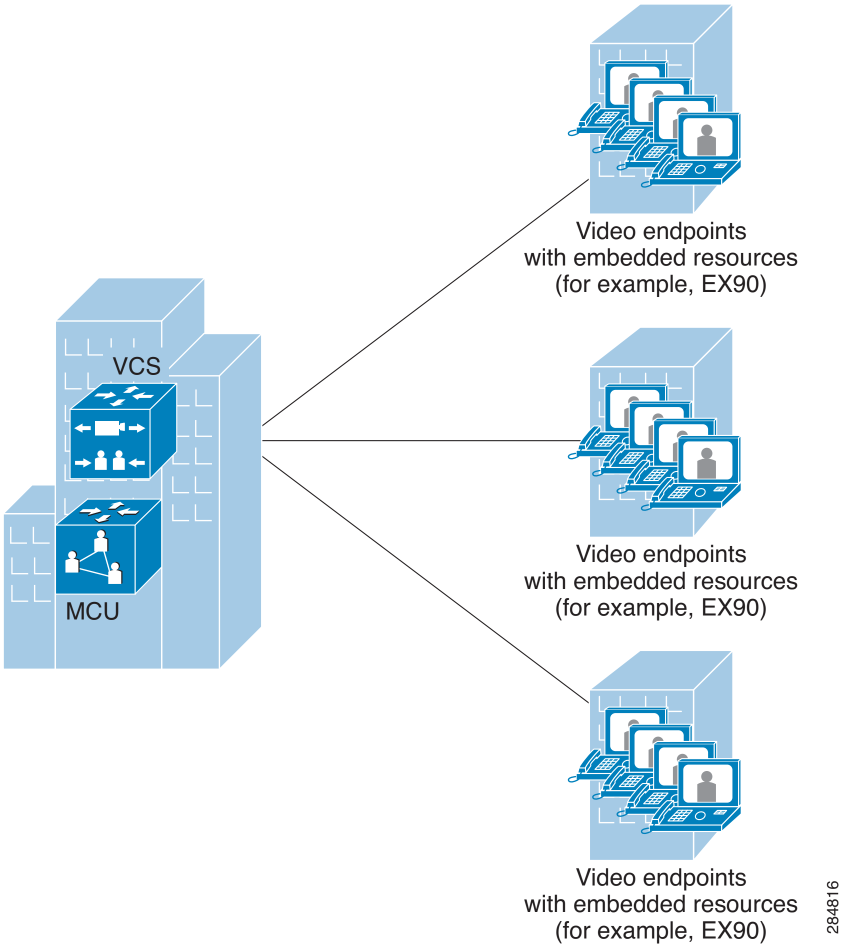
Figure 7-7 Hub and Spoke Network with Centralized Dedicated Video Resources (No Embedded Resources)


VCS
MCU
Video endpoints without embedded resources
Video endpoints without embedded resources
Video endpoints without embedded resources
284815

In general, centralized video resource allocation architectures are better suited for scenarios where not many remote endpoints exist and their remote location and bandwidth usage does not induce undesirable effects on the media streams to be transported in the WAN links.

A modification to the previous scenario is presented in the hybrid example in Figure 7-8, in which embedded video resources are located at the remote sites. In this scenario, the centralized dedicated video resources would be utilized only when a video conference involves a video endpoint without embedded resources in the central location or when the number of participants in the video conference exceeds the capacity that the embedded video resources can handle.

Figure 7-8 Hub and Spoke Network with Centralized Dedicated Video Resources and Embedded Video Resources at the Remote Sites



Many other scenarios are possible, and therefore different strategies and/or restrictions can also apply to the centralized video resource approach.

Distributed Video Resource Allocation

Distributing the dedicated video resources throughout various locations has several advantages; chief among them are the WAN link bandwidth savings and less likelihood of inducing undesirable effects on the media streams (since many of them are locally terminated). However, a distributed allocation also has some limitations. For example, certain video calls still have to traverse the WAN links, and these streams are still limited by the WAN link properties.

Cost effectiveness of the solution can be maximized by considering the following points when determining whether or not to deploy distributed dedicated video resources:

- What is the expected video call utilization pattern at the location(s) of the video resources?
- Can the current bandwidth of the WAN link(s) support the expected usage pattern(s) of the remote location(s) without inducing undesirable effects in the video streams?
- Will the limitations or effects (if any) of the media transmissions over the WAN links be acceptable for the intended use cases?
- Would using distributed embedded video resources satisfy the planned use cases?
- Will the video solution be able to grow adequately without distributed dedicated video resources, given the current and planned network topologies?

Additionally, when using a distributed allocation for dedicated video resources, it is important to understand the bridge selection algorithm of the call control elements in order to decide where best to locate the video resources. For instance, the dedicated video resources might be reserved based on the time zones of the endpoints and the resources. Other alternatives involve video resource reservation based on video location or manual reservation. In any case, the importance of understanding the selection algorithm derives from the need to understand where the streams are ultimately terminated in order to distribute the video resources more efficiently.

Creating a Video-Ready Network

Video brings significant benefits for businesses, such as superior collaboration, lower travel costs, and personalized advertisements. However, video applications also introduce additional challenges for the underlying network infrastructure and IT departments. For instance, how does one configure the network for video? How should IT departments prioritize and scale video? How do they protect other applications from being swamped by high-bandwidth video streams? To support these enterprise video applications, a tightly controlled network foundation providing the following services is required:

- Optimized Video Delivery, page 7-14
- Security of Video Applications, page 7-16
- Scalability and Performance, page 7-16

Optimized Video Delivery

For video to be an efficient collaboration tool, the user experience must be of high quality. To ensure the user experience quality, the video delivery must be optimized to meet the organization's requirements. The following sections offer guidelines on how to optimize the video delivery:

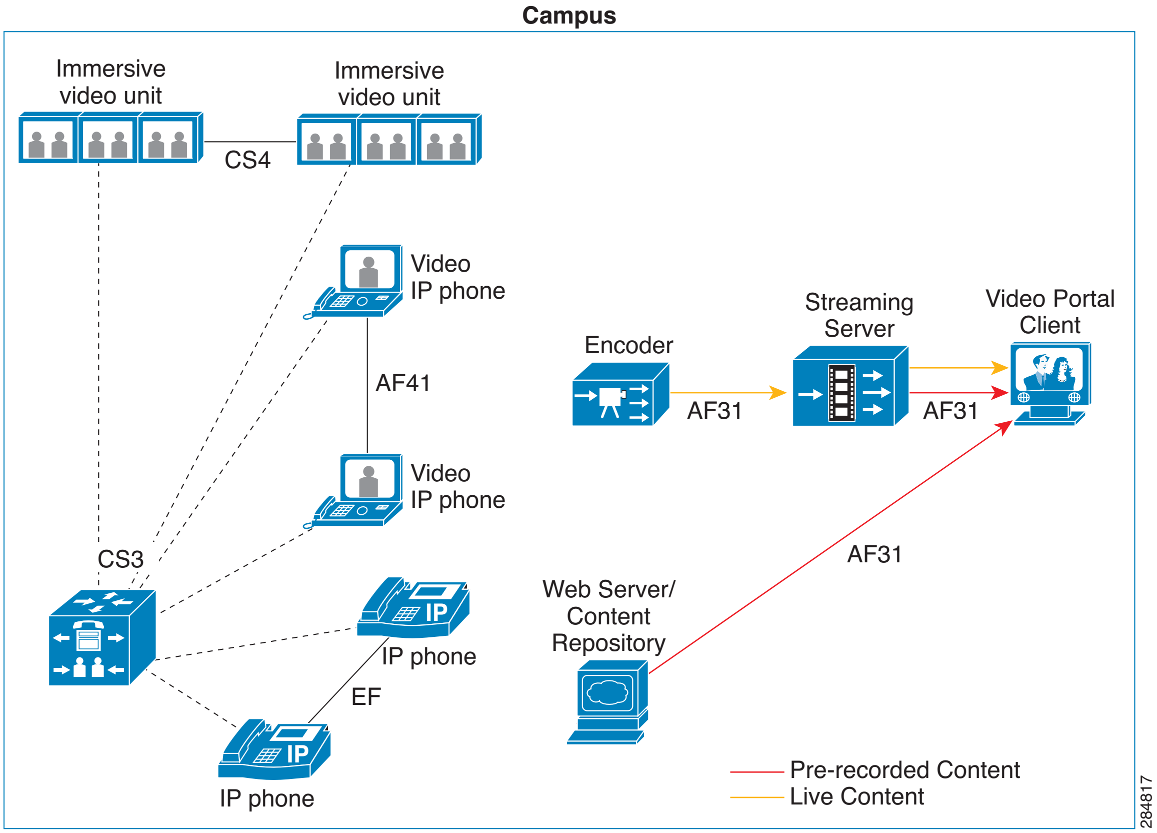
- Quality of Service (QoS), page 7-14
- Content Sharing Technologies, page 7-15
- Reliability, page 7-16

Quality of Service (QoS)

The first step in optimizing the delivery of video is to identify the traffic of interest and to apply a differentiated Quality of Service (QoS). QoS helps the organization to provide application intelligence to differentiate between business-critical and noncritical video streams, and to keep the latency, jitter, and loss for selected traffic types within an acceptable range. Furthermore, priority queuing should be used over other queuing mechanisms whenever possible to provide a better video experience. In the case of multiple video applications on converged networks (TelePresence and IP video telephony combined in the same network), Cisco recommends differentiation of QoS classes per application.

Figure 7-9 depicts an example of multiple IP video applications and IP voice converging in the same network. In this example, immersive video, videoconferencing, video-on-demand, and voice over IP are identified and assigned the recommended QoS markings to provide the required service level, thus avoiding over-provisioning or overlap of applications in the queues.

Figure 7-9 Recommended QoS Traffic Markings in a Converged Network



For more information about QoS in video solutions, refer to the chapter on Quality of Service and Call Admission Control, page 5-1.

Content Sharing Technologies

Depending on the endpoints being deployed as part of the video solution, it is important to consider the content sharing standards supported and required by the video endpoints and how they will converge or interoperate if necessary. There are currently three main content sharing technologies used by IP video solutions: Binary Floor Control Protocol (BFCP), H.239, and auto-collaborate. H.323 endpoints make use of the H.239 standard to provide the content sharing functionality, while SIP endpoints may use auto-collaborate or the newer standard BFCP.

For information about presentation sharing and content sharing technologies, refer to the *Cisco TelePresence Interoperability Deployment Guide*, available at

http://www.cisco.com/en/US/products/ps8332/products_device_support_tables_list.html

Reliability

The performance and availability of a video-ready campus must be monitored pro-actively and measured across the network. Alternate paths in the case of failure should also be offered to ensure the reliability of the video solution. For information about how to design highly available networks, refer to the *Campus Network for High Availability Design Guide*, available at

http://www.cisco.com/en/US/solutions/ns340/ns414/ns742/ns815/landing_cHi_availability.html

Security of Video Applications

Whenever possible, a video-ready network should integrate video security to protect against unauthorized access to video applications. Mitigation of attacks and protection of traffic from snooping and intrusion by malicious users is essential, and so is preventing malicious users from transmitting unauthorized video. A variety of techniques can be used to secure a video network, from network virtualization techniques to segregate video traffic, to the use of Trusted Relay Points (TRP) for software clients residing in the data VLAN or Session Border Controllers (SBC) for topology hiding from the exterior world. For information about how to secure video networks, see the chapter on Security for Video Communications, page 9-1.

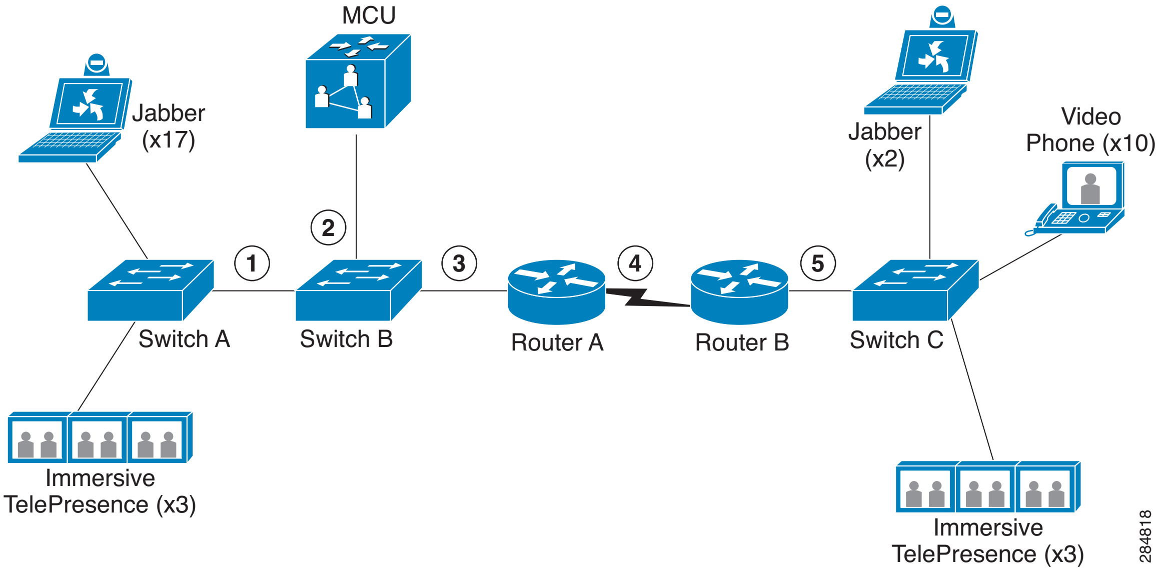
Scalability and Performance

Network scalability is critical to supporting increasing bandwidth demands as more video applications or video users are deployed. To maintain optimal performance, the network should easily accommodate higher bandwidths, scaling to support high-definition (HD) video streams and in some cases even multiple HD video streams simultaneously. Therefore, it is crucial to size the network adequately for the expected traffic generated by the video applications to be deployed.

The first step in sizing the network is to understand the endpoint and user requirements. Next, determine how the media flows will behave during conferences and point-to-point calls. Then add the considerations for voice and data traffic and backup measures required for reliability. Figure 7-10 illustrates an example scenario where 20 immersive and desktop video endpoints are located in the headquarters campus while 15 miscellaneous endpoints are located in the branch. The expected usage pattern in this example is as follows:

- Headquarters desktop endpoint users and immersive endpoint users will generate peak usage of 7 calls among each other.
- Headquarters endpoints will generate and/or receive peak usage of 6 calls to/from the branch office.
- Calls on headquarters desktop endpoints use 1.3 Mbps while the immersive endpoints use 12 Mbps.
- Branch video IP phones use 1 Mbps, desktop endpoints use 1.3 Mbps, and immersive endpoints use 12 Mbps.
- At peak times, branch users will generate a maximum 4 calls among each other and generate or receive 6 calls (as listed above) to/from the headquarters.
- Headquarters and branch users need to access 10 Mbps (combined) of data applications between sites.

Figure 7-10 Determining Capacity Requirements for a Video-Ready Network



The above requirements are obviously simplistic. A very complex network and deployment would have a longer set of requirements and applications to support. But with the list of requirements in the example above, we can determine the following:

- The maximum expected bandwidth usage is 49 Mbps for the video streams in the uplink between switch A and switch B (link 1 in Figure 7-10), assuming the worst-case scenario of:
  - Seven participants in a local multipoint call of the desktop video endpoints:

    7 ∗ (1.3 Mbps) = 9.1 Mbps
  - Six participants in a multipoint call of the immersive endpoints, 3 local and 3 remote:

    3 ∗ (12 Mbps) = 36 Mbps
  - Six participants in a multipoint call of the local desktop endpoints, 3 local and 3 remote video IP phones:

    3 ∗ (1.3 Mbps) = 3.9 Mbps

Note This bandwidth calculation does not consider the extra bandwidth for data applications or any other extra bandwidth that is necessary (for call signaling, for example). Before dimensioning the LAN, you need to include these other bandwidth requirements.

- In a worst-case scenario, the branch WAN link (link 5 in Figure 7-10, but the same considerations apply for links 3, 4, and 6) would use 43.6 Mbps of bandwidth for video streams, assuming:
  - Four participants in a multipoint call between 4 video phones. Because no local Multipoint Control Unit (MCU) is available, all the streams have to travel to the headquarters for the multipoint call to occur.

    4 * (1 Mbps) = 4 Mbps
  - Six participants in a multipoint call between 5 desktop video endpoints and 1 video phone, 3 local endpoints (2 desktop video endpoints and 1 video phone) and 3 remote endpoints:

    2 * (1.3 Mbps) + 1 Mbps = 3.6 Mbps
  - Six participants in a multipoint call on the immersive endpoints, 3 local and 3 remote:

    3 * (12 Mbps) = 36 Mbps

  Note This bandwidth calculation does not consider the extra 10 Mbps requested for data applications or any other extra bandwidth that is necessary (for call signaling, for example). These requirements also need to be added as part of the sizing process.

- Finally, for link 2 servicing the MCU in Figure 7-10, we can anticipate that 92.6 Mbps of video streams will traverse it at its peak if we consider the two previous bandwidth calculations:

  49 Mbps + 43.6 Mbps = 92.6 Mbps

Note The general rule that has been thoroughly tested and widely used is to over-provision video bandwidth by 20% in order to accommodate a 10% burst and the Layer 2-to- Layer 4 network overhead. Furthermore, the above calculations are based on the worst-case scenario for the usage patterns provided in the example, but they do not consider the case where all video users want to make video calls at the same time (known as 100% call completion).

In summary, the network capacity and performance design in a video-ready network should allow for video forwarding without introducing significant latency of the call completion rate desired or the usage patterns expected. Refer to your endpoint documentation to determine the amount of bandwidth required per call.

Integration with Standalone Video Networks

Whether replacing a previous video network solution or trying to converge the video network solution under the same call processing platform elements, integration could pose some challenges for the IT department. Understanding the options and guidelines for integration will provide a better experience for the integrator and the user.

The integration approach differs depending upon the call signaling protocol used. The following sections outline the general guidelines for the two most widely used protocols in video networks today:

- Integration with Standalone H.323 Video Networks, page 7-19
- Integration with Standalone SIP Video Networks, page 7-20

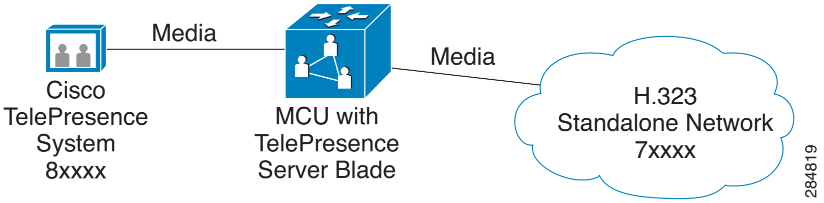
Integration with Standalone H.323 Video Networks

H.323 is a very well define protocol, which makes interoperability with H.323 call processing elements considerably easier than with multi-vendor SIP elements; however, H.323 is not as service-rich as its counterpart SIP. For example, Cisco has implemented the ability to switch screens depending on who the active speaker is (smart switching), but this feature is not natively available in H.323 networks.

Whenever possible, use the native interoperability of the video endpoints to connect them directly to the H.323 network, provided that this does not cause the loss of required features such as smart switching. Otherwise, if feature retention is critical or if interoperability cannot be obtained point-to-point natively, then connect the H.323 endpoints through a video transcoder or an interoperability-enabled conference bridge. A non-overlapping dial plan is also recommended, and different access codes can be used between the networks to indicate to the call processing agents that a hop to the next video system is required to complete the call.

Figure 7-11 illustrates the use of a Multipoint Control Unit (MCU) to connect a Cisco TelePresence System to a third-party H.323 standalone video network.

Figure 7-11 Integrating Standalone H.323 Networks


Media
Cisco TelePresence System 8xxxx
MCU with TelePresence Server Blade
Media
H.323 Standalone Network 7xxxx
284819

Integration with Standalone SIP Video Networks

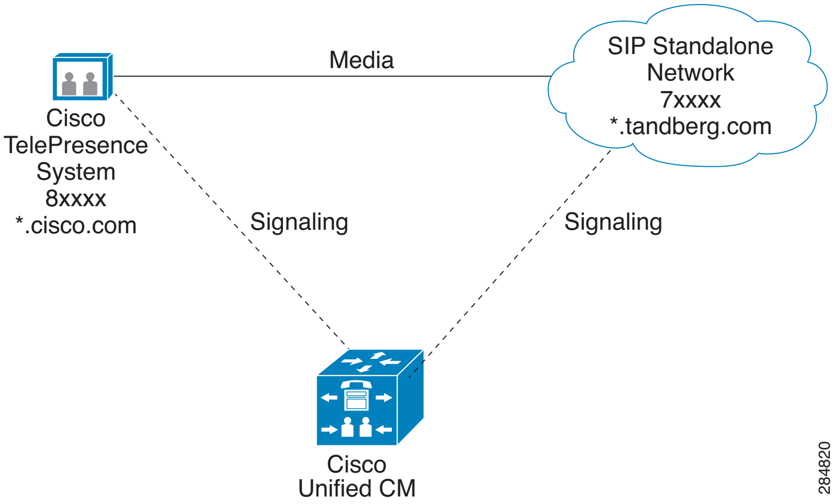
SIP video networks are more feature-rich than H.323 networks and can enable very useful features when all the endpoints support the. However, SIP is not as well defined as H.323, thus making interoperability with it more challenging.

If an endpoint conforms tightly to the SIP standard, then call agents can make use of the native video interoperability features available within the call processing agents. Otherwise, video networks can be bridged to each other with video transcoders or interoperability-ready multipoint control units.

Cisco recommends that you do not use an overlapping dial plan between the video networks, but do use an access code to indicate to the call processing agents to route between video networks and avoid inter-digit time-out. If uniform resource identifier (URI) dialing is used for the dial plan, Cisco recommends using different domains for ease of administration.

Figure 7-12 illustrates integration of a Cisco TelePresence System with a SIP standards-based third-party system. This example uses native interoperability along with different domains for dialing.

Figure 7-12 Integrating Standalone SIP Networks


Media
SIP Standalone Network 7xxxx *.tandberg.com
Cisco TelePresence System 8xxxx *.cisco.com
Signaling
Signaling
Cisco Unified CM
284820

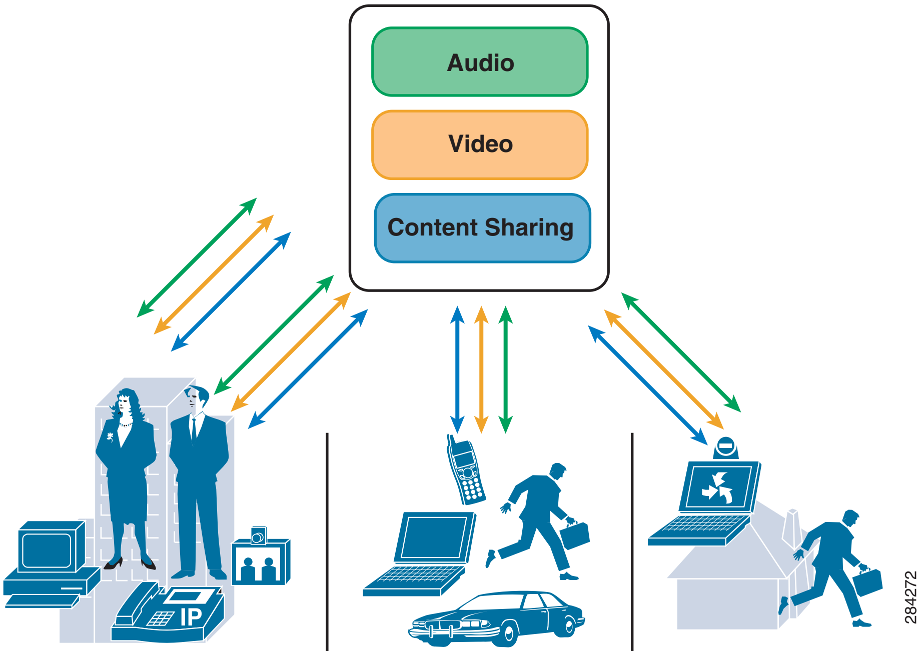
CHAPTER 8

Collaborative Conferencing

Revised: March 30, 2012, OL-27011-01

When there are three or more participants involved in a call, the call becomes a conference. In collaborative conferencing, the audio, video and content from some or all of the attendees in a meeting is mixed into a single stream and the stream is sent back to the attendees. The processing of audio and video is performed by a Multipoint Control Unit (MCU) or some multipoint devices. Figure 8-1 illustrates a conference involving both internal and external participants, mobile and remote workers, or even attendees from different organizations.

Figure 8-1 Conceptual View of Collaborative Conferencing



Conference Type

Most conferencing products support two types of conference, ad-hoc and scheduled. Scheduled conferences require a special tool such as Cisco TelePresence Management Suite (TMS), for example, that integrate with the calendar applications such as Microsoft Exchange or IBM Domino to provide the scheduling functionality.

Ad-hoc Conference

In ad-hoc conferences, the conference initiator creates the conference and invites participants to join without sending any prior notification about the conference information. Typically, the conference initiator creates the conference by pressing the Conference key on the endpoint and calls the participants to add them to the conference. Endpoints that do not have the Conference key cannot initiate the conference but can be invited to join the conference. Resources for an ad-hoc conference cannot be reserved. The conference can be created only if enough resources are available at the conference creation time. In this case, the call processing agent (Cisco Unified Communications Manager, for example) controls the conference.

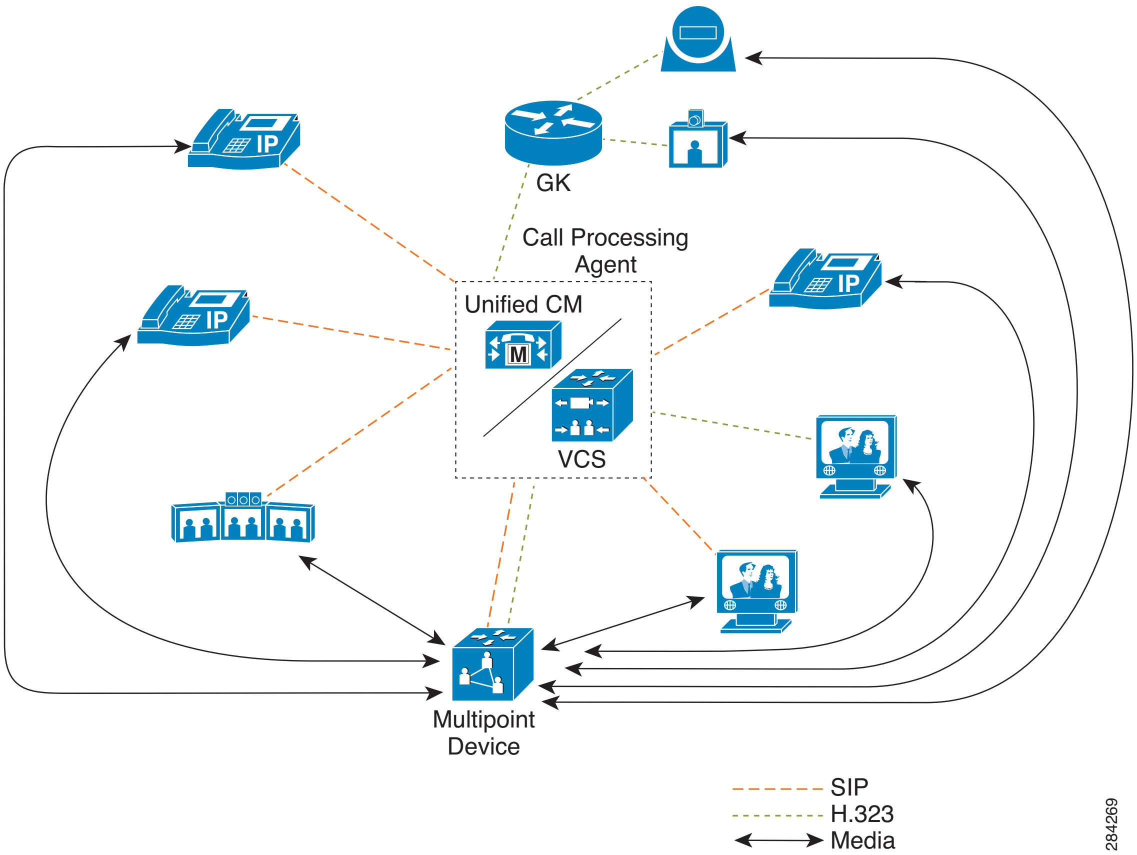
Scheduled Conference

In this type of conference, the conference organizer schedules the conference using the product scheduling tool or calendar application that integrated with the conference product scheduling function. The conference resources are reserved at the time the conference is scheduled. The conference cannot be scheduled if there are insufficient resources. To join the conference, attendees can dial into the conference directly or click on the conference link inside the meeting invitation and have the system call the attendees. In this case, the multipoint device controls the conference and the call processing agent routes the dialed call to the correct multipoint device.

Conferencing Infrastructure

As shown in Figure 8-2, the call processing agent and multipoint device are the main components in the conferencing infrastructure. The call processing agent handles the signal and controls the call on the endpoints. In the case of a scheduled conference, the call processing agent routes the call to the multipoint device. Cisco Unified Communications Manager (Unified CM) and the Cisco TelePresence Video Communication Server (VCS) are the most common call processing agents used in the Cisco video conference products. Unified CM can interconnect with VCS through a SIP trunk so that endpoints registered to one agent can call or conference in endpoints registered to the other agent.

Figure 8-2 Conferencing Components



The main function of the multipoint device is to handle the media (audio and video) from the endpoints during the conference. Typically, the connection between the call processing agent and the multipoint device is SIP or H.323.

Conference Multipoint Device

The multipoint device is the central component in the conferencing infrastructure and it can be hardware or software based. The multipoint device processes video streams from all participating endpoints and transmits a composite image of all participants back to the originating endpoint devices. This composite view enables all participants to see each other simultaneously. The continuous presence view can display multiple windows (participants) in a variety of layouts. Each layout offers the ability to make one of the windows voice-activated, which is useful if there are more participants in the conference than there are windows to display them all in the composite view.

Transcoding and Switching

The multipoint platform can be based upon either a transcoding or switching architecture. Both architectures provide advantages that need to be considered when selecting a multipoint platform for deployment.

- **Transcoding** involves specialized video hardware that decodes the incoming video stream and then re-encodes it before sending it on. For example, the Cisco TelePresence Server, Cisco Media Experience Engine (MXE), and Cisco Multipoint Control Unit (MCU) all use the transcoding architecture.
- **Switching** does not require specialized video hardware but uses software instead. The incoming video and audio streams are copied and redirected to the correct endpoints in the conference, with no manipulation of the video stream. For example, the Cisco TelePresence Multipoint Switch uses the switching architecture.

Table 1 describes the advantages and disadvantages of the transcoding and switching architectures of the multipoint platform.

Table 8-1 Comparison of Transcoding and Switching

Architecture	Advantages	Disadvantages
Transcoding	• Active presence support[1] • Ability for endpoints to connect at different bandwidth speeds and resolutions • Often endpoints with Far End Camera Control (FECC) can customize layouts • Ability to scale video (translate) between endpoints • Supports Telepresence Interoperability Protocol (TIP) and standards-based endpoints	• Latency is introduced due to decoding and re-encoding video • Higher cost per port • Typically harder to scale
Switching	• Latency is low (less than 10 ms) • Lower cost per port	• Limited to basic full-screen video switching (No active presence; only one site is visible on each screen) • Endpoints must support and agree on a single resolution and frame rate[2]

[1] Active presence provides a view of multiple participants on a single display, with the active participant shown in a full screen while the other participants appear as a picture-in-picture overlaid along the bottom of the display.

[2] Beginning with Cisco TelePresence System Release 1.7, Cisco TelePresence System devices using Smart Media can adjust the video resolution based on the IP network conditions.

Multipoint Deployment Guidelines

There are two options for deploying any multipoint technology, namely centralized and distributed. It is important to think through the deployment strategy to optimize the user experience, localize resources, and ensure reliable multipoint calls. Consider the followings when designing your multipoint solution deployment:

- The number of endpoints, which in turn determines the number of multipoint devices required
- The geographic location of the endpoints

Together these factors determine the multipoint deployment option (centralized or distributed), the number of multipoint devices, and the physical location of the multipoint devices.

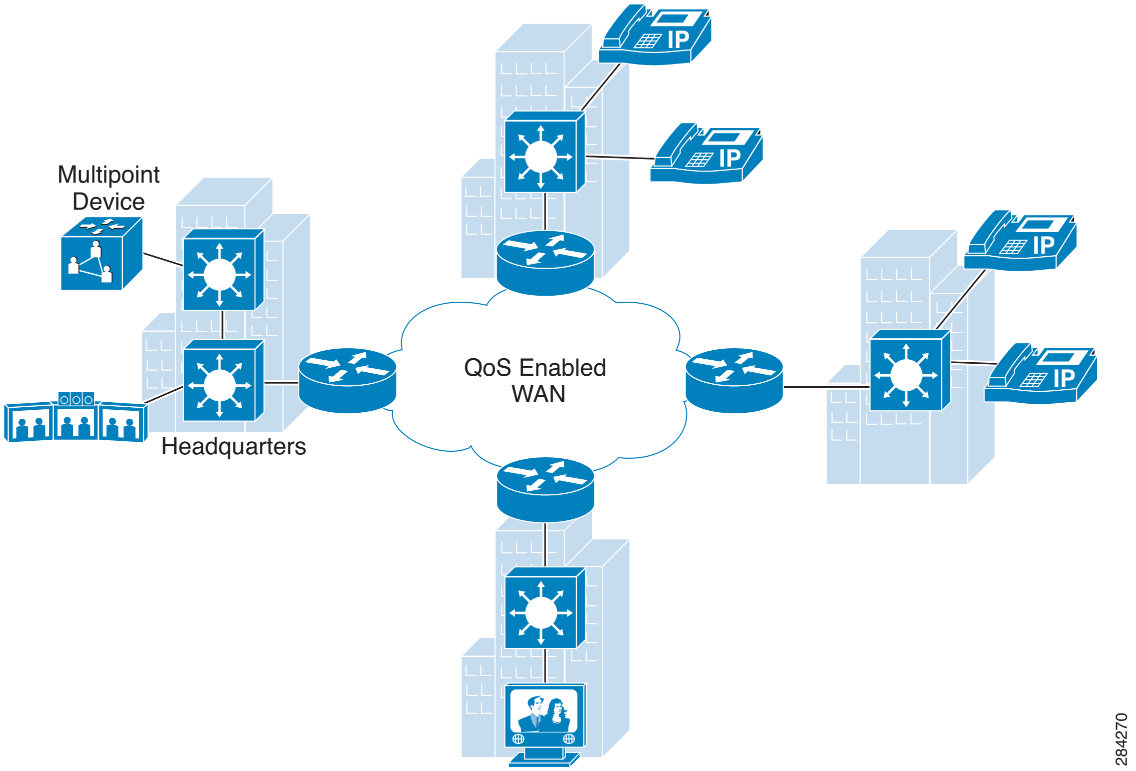
Centralized Deployment

Centralized designs are recommended for multipoint device deployments with a small number of endpoints, or for larger deployments that cover a limited geographic area.

In centralized deployments, the multipoint device can be located in a regional or headquarters campus site with the necessary WAN bandwidth available to each of the remote sites (as well as the necessary LAN bandwidth within the campus). Cisco recommends that you locate the multipoint device centrally based on the geographic location of the endpoints. (However, this might not be possible in all network layouts.) Centrally locating the multipoint device prevents unnecessary latency caused by backhauling calls to a site at the far edge of the network.

Figure 8-3 illustrates a deployment with a small number of endpoints that uses the multipoint device placed centrally in the headquarters to minimize latency for multipoint meetings.

Figure 8-3 Centralized Multipoint Deployment



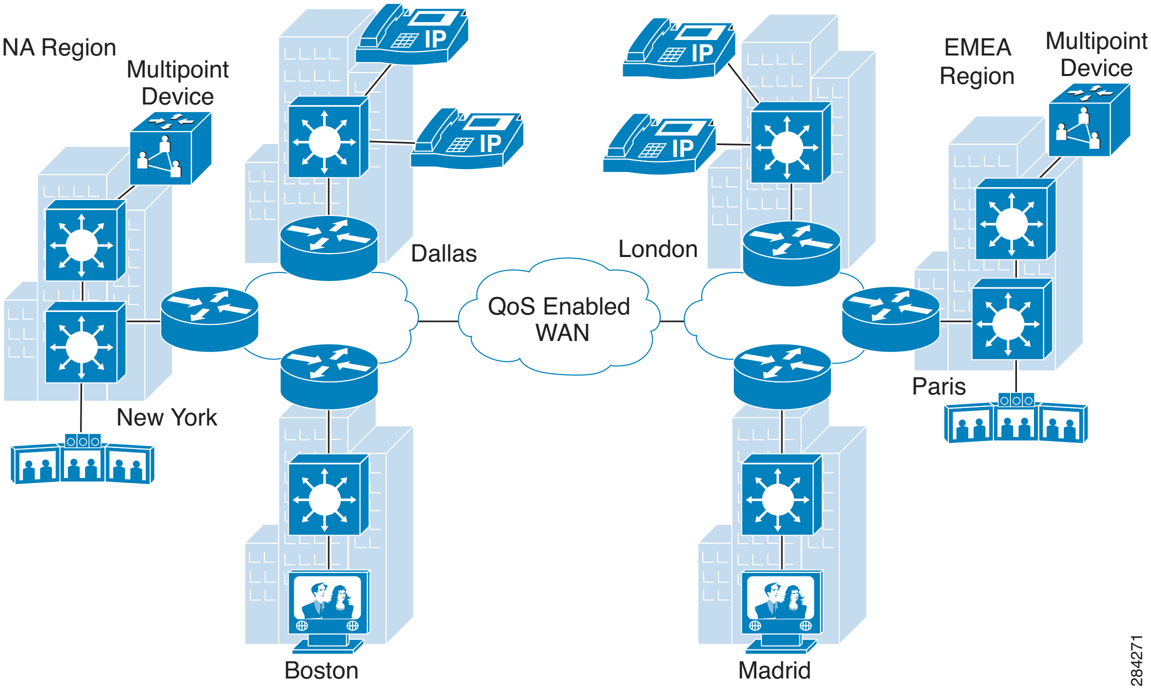
The multipoint device should be located at a site that provides network latency that adheres to the solution requirements. Also, the site should be provisioned with adequate bandwidth for the number of endpoints deployed on the network. Bandwidth requirements vary depending on the desired maximum call rate of endpoints and the number of endpoints connecting to the multipoint device. Provision based on the maximum bandwidth that a particular endpoint requires for the desired rate and resolution. For details on network latency and bandwidth requirements, refer to the respective solution deployment or design guide available at http://www.cisco.com.

Distributed Deployment

Cisco recommends the distributed configuration for large deployments or deployments with a small number of endpoints in separate geographical regions. As the network grows, it is very advantageous to localize multipoint devices to minimize latency and save bandwidth.

Figure 8-4 illustrates the deployment where the multipoint devices are placed centrally in each region (NA and EMEA) but are distributed globally.

Figure 8-4 Distributed Multipoint Deployment



In the distributed environment, multipoint devices should be located at sites that provide network latency adhering to solution requirements and adequate bandwidth for the number of endpoints supported by each site.

The following additional guidelines should be observed when deploying multipoint devices:

- Build a resilient multipoint solution architecture using multiple multipoint devices. In case one of the multipoint devices fails, conferences can still be started from the operational device.
- Cascade multipoint devices to support a larger conference deployment and to minimize bandwidth consumption. Cisco TelePresence Conductor can be used to cascade multipoint devices.
- Use Cisco TelePresence Conductor to preferentially select a multipoint device for conferences based on their properties (for example, geographic location or video quality).
- Choose a deployment that allows you to build a scalable multipoint solution. For example, if you have endpoints deployed in multiple geographical regions and expect the number of endpoints to grow in each regions, use the distributed multipoint deployment
- Configure a solution so that traffic is load-balanced to the multipoint devices in order to achieve maximum resource utilization.
- Cisco recommends using the scheduling option for larger deployments. This would simplify the conference creation.
- Wherever possible, Cisco recommends dual-registering (SIP and H.323) the MCU in mixed environments so that endpoints can connect in their native protocol without interworking.

Multipoint Solution Selection

When deciding on the multipoint solution suitable for your deployment, you need to consider several factors (such as your organization's current deployment and the intended endpoint combinations) and to decide which multipoint features have priority. Some common factors to consider when designing the multipoint solution architecture include:

- Active presence
- Scalability
- Network latency
- Endpoints support
- Multi-screen compatibility
- Collaboration (content sharing)
- Scheduling

Multipoint Selection Based on Endpoints

When deciding on a multipoint device, start with your endpoints. In some cases, a combination of endpoints could help eliminate or determine the multipoint device option for the deployment, as in the following scenarios:

- In a deployment with heterogeneous endpoints that support different video formats, video codecs, resolutions, or frame rates, the best option would be multipoint devices that can do both transcoding and transrating. In this case, the Cisco TelePresence Multipoint Switch would not be a good choice because it is based on a switching architecture.
- A deployment that contains both Telepresence Interoperability Protocol (TIP) and standards-based endpoints would require a video gateway to inter-operate between the protocols. For example, if the conference environment has a Cisco TelePresence Multipoint Switch and a mix of Cisco TelePresence System 3200 Series, Cisco IP Video Phone E20, and Cisco TelePresence System EX90 endpoints registered to Cisco Unified CM, the Cisco Media Experience Engine (MXE) might be the obvious option to add (from a budget perspective) because it allows interoperability between endpoints.

However, in most cases there can be more than one option, so you should also consider the features that you require.

Multipoint Selection Based on Features

The multipoint solution options can be further narrowed down by focusing on which features have the highest priority for your conferencing deployment. The following partial list represents features that affect the decision of which multipoint solution to choose:

- Support for Telepresence Interoperability Protocol (TIP), standards-based, or third-party endpoints
- Multi-screen or single-screen support
- Support for One Button To Push (OBTP)
- Active presence or continuous presence
- Content sharing
- H.239, Binary Floor Control Protocol (BFCP), or Auto Collaborate content sharing translation
- WebEx OneTouch integration
- Layout changes possible from endpoints
- Switching policy support (room versus speaker switching)
- SIP and H.323 support

For a complete list of features supported in each multipoint product, refer to the respective product data sheet available at http://www.cisco.com.

Conference Resource Capacity Planning

There are several factors involved in determining the types and number of video conferences that the multipoint device can support. These sizing factors are different for different multipoint products. Multipoint devices can also support higher capacity when using standard definition (SD) as compared to high definition (HD) for video conference.

Conference capacity depends on the following factors:

- Type of resolution or video format required for the video conference
- Total number of ports that the multipoint device can support
- Number of ports that the multipoint device can dedicate to each protocol
- Whether the multipoint devices are cascaded

If the video conference utilizes a Cisco High-Density Packet Voice Video Digital Signal Processor Module (PVDM3) or later model of digital signal processor (DSP) on a Cisco Integrated Services Router (ISR) as the bridge, use the DSP calculator in the Cisco Unified Communications Sizing Tool to determine the conference capacity. The Cisco Unified Communications Sizing Tool (Unified CST) is available to Cisco employees and partners with proper login authentication at http://tools.cisco.com/cucst.

CHAPTER 9

Security for Video Communications

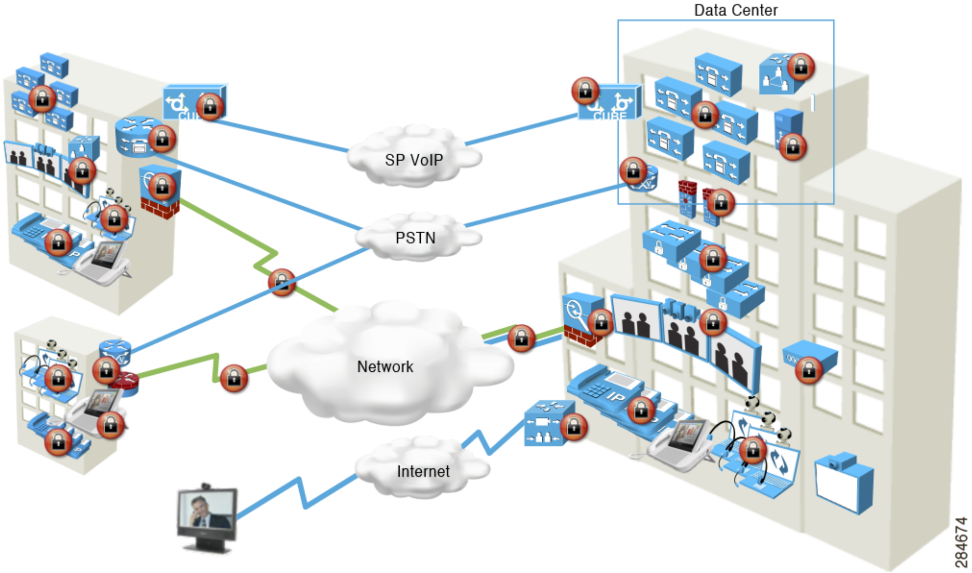
Revised: March 30, 2012, OL-27011-01

Securing video communications over the IP network in an enterprise requires implementing security for both the Unified Communications components and the network infrastructure that those communication streams traverse. This chapter focuses on the design and implementation options available within the Cisco Unified Communications System and the Cisco TelePresence Solution for securing the integrity, reliability, and confidentiality of video calls within an enterprise IP Telephony network. For more information on data network security, refer to the Cisco SAFE Blueprint documentation available at

http://www.cisco.com/en/US/netsol/ns744/networking_solutions_program_home.html

To securely implementing voice and video communications, Cisco recommends creating security policies associated with every network technology deployed within an enterprise (see Figure 9-1). The security policy defines which data in your network is sensitive so that it can be protected properly when transported throughout the network. Having a security policy helps define the security levels required for the types of data traffic that are on a network.

Figure 9-1 Security and Hardening Options in a Cisco Unified Communication System



Hardening the Cisco Unified Communications network involves establishing and maintaining authenticated communication streams, digitally signing configuration files, and encrypting media streams and call signaling between various Cisco Unified Communications and Cisco TelePresence components. All of these security features are not required for every network, but they provide options for increasing levels of security.

This chapter provides the design guidelines for these features. For the product configuration details, refer to the following security documentation for your specific version of Cisco Unified Communications Manager (Unified CM) and Cisco TelePresence:

- *Cisco Unified Communications Manager Security Guide*

  http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html
- *Cisco Unified Communications System SRND*

  http://www.cisco.com/go/ucsrnd
- *Cisco TelePresence Design Guide*

  http://www.cisco.com/en/US/solutions/ns340/ns414/ns742/ns819/landing_vid_tPresence.html

Network Infrastructure Security

Securing video communications requires securing the network that is used for transporting the calls. This can be achieved by building layers of security, starting at the access port, continuing across the network and to the Internet edge. Cisco recommends always using firewalls, access control lists, authentication services, and other Cisco security tools to help protect your network infrastructure devices from unauthorized access.

Restricting access to the network devices is one of the most important requirements in securing the infrastructure. A typical enterprise network consists of many components, including routers, switches, firewalls, and intrusion prevention systems. Attackers are constantly trying to access these devices on networks. Restricting access to the management interface of each device lowers the opportunities that attackers have to compromise them. All of the devices on a network should be secured appropriately. Administrative and operational management of the network devices should be done using secured protocols such as Secure Shell (SSH) and Hyper-Text Transfer Protocol Secure (HTTPS). Transmission of passwords and configuration information over clear text, used in protocols such as Telnet, should be avoided as much as possible.

In addition to securing access to the infrastructure, the services used in the operation of a network also need to be secured. These include Domain Name System (DNS), Network Time Protocol (NTP), Dynamic Host Configuration Protocol (DHCP), and signaling protocols such as Session Initiation Protocol (SIP) and H.323. These services, which are vital to the successful operation of a network, are also prime targets for an attacker. Disrupting any of these services can cause denial of service and availability problems for the Unified Communications systems.

Separate Auxiliary VLAN

Cisco recommends implementing separate VLANs for RTP traffic (voice and video) and data traffic in a Unified Communications environment. In this configuration, all Cisco IP Phones and TelePresence endpoints are placed in a voice VLAN that is separate from the data VLANs. This implementation provides the following benefits:

- It makes it convenient to design VLAN access control lists (VACLs) that can be used to restrict traffic between voice and data network components. This also allows network administrators to more effectively implement management access restrictions on the network.
- It provides address space conservation and voice device protection from external networks. Private addressing of phones on the voice or auxiliary VLAN ensures address conservation and prevents phones from being accessible directly through public networks.
- It enables simplified Quality of Service (QoS) configuration and management. It also allows the QoS trust boundaries to be extended to voice and video devices without extending the trust and the QoS features to PCs and other data devices.
- VLAN access control, 802.1Q, and 802.1p tagging can prevent attempts by data devices to spoof information and gain access to priority queues through packet tagging.

Note The Cisco Unified IP Phones and the Cisco TelePresence endpoints have different service requirements, and placing them together in a single VLAN makes it more complicated to design regular access control lists.

Device Security

Cisco Unified IP Phones and TelePresence endpoints have multiple configuration options for securing them against attacks. However, these devices should not be considered hardened by default at the time of initial configuration. The security features vary among the different endpoints and include:

- Secure Management over HTTPS and SSH, page 9-3
- Administrative Passwords, page 9-4
- Device Access, page 9-4
- Signaling and Media Encryption, page 9-4

Refer to the endpoint administration guides for information on configuration details for these features. Also, refer to the phone hardening information in the *Cisco Unified Communications Manager Security Guide*, available at

http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

Secure Management over HTTPS and SSH

Cisco TelePresence endpoints support management through Secure Shell (SSH) and Hyper-Text Transfer Protocol over Secure Sockets Layer (HTTPs). Access to the endpoints using HTTP, HTTPS, SSH, or Telnet can be configured in the Network Services setting on the endpoint itself.

Cisco Unified IP Phones can be restricted to use HTTPS only or enabled for both HTTP and HTTPS.

Administrative Passwords

The endpoints ship with default administrative passwords, and Cisco recommends changing the passwords at the time of installation. Access to management functions should be restricted to authorized users with administrative privileges.

Device Access

The endpoints can be assigned to users who are given access based on defined roles and privileges. Passwords and PINs can be specified for these users to enable SSH or Telnet and web-based access. A credential management policy should be implemented to expire and change passwords periodically and to time-out logins when idle. This is necessary for limiting access to the devices to verified users.

For information on user authentication and credential management configurations, refer to the following documentation:

- *Cisco Unified Communication Manager Administration Guide*

  http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html
- *Securing Cisco TelePresence Products*

  http://www.cisco.com/en/US/products/ps8332/products_installation_and_configuration_guides_list.html

Signaling and Media Encryption

For supported Cisco Unified Communications devices, signaling and media can be encrypted to prevent eavesdropping and reconnaissance attacks on active calls and during call establishment. The protocols and mechanisms used to provide secure communications and signaling within Unified Communications deployments include the following:

- Transport Layer Security (TLS), page 9-4, used for encrypting signaling traffic
- Secure Real-Time Transport Protocol (SRTP) and Secure Real-Time Transport Control Protocol (SRTCP), page 9-5, used for encrypting media
- Datagram Transport Layer Security (DTLS) Secure Real-Time Transport Protocol (SRTP), page 9-5, used for SRTP master key negotiation and/or exchange
- Digital Certificates, page 9-5
- Certificate Authority Proxy Function (CAPF), page 9-6
- The Certificate Trust List (CTL), page 9-6

Transport Layer Security (TLS)

Transport Layer Security (TLS) is a protocol designed to provide authentication, data integrity, and confidentiality for communications between two applications. TLS is based on Secure Sockets Layer (SSL) Version 3.0, although the two protocols are not compatible. The latest version, TLS 1.2, is defined in IETF RFC 5246. TLS operates in a client/server mode, with one side acting as the server and the other side acting as the client. TLS uses a handshake protocol to allow the client and server to authenticate each other using public key cryptography (digital certificates). This also enables reliable negotiation of a compression algorithm, message authentication algorithm, encryption algorithm, and the necessary cryptographic keys before any application data is sent.

Data authentication and encryption of the SIP signaling between Cisco Unified CM, Cisco Unified IP Phones, and Cisco TelePresence System components, is implemented using the Transport Layer Security (TLS) protocol. TLS is also used for the authentication and confidentiality of the web services signaling between the various Cisco TelePresence components.

The encryption of the signaling protocol is done using the Advanced Encryption Standard (AES) algorithm, using symmetric keying. Message authentication is done with the HMAC-SHA1 hash algorithm. The negotiation of keying material is done securely within the TLS Handshake Protocol layer through the Client and Server Key Exchange messages.

Secure Real-Time Transport Protocol (SRTP) and Secure Real-Time Transport Control Protocol (SRTCP)

Data authentication and confidentiality of the Real-time Transport Protocol (RTP) voice and video media flows use Secure Real-time Transport Protocol (SRTP) for both point-to-point and multipoint TelePresence meetings.

Secure RTP (SRTP) and Secure Real-time Transport Control Protocol (SRTCP) are both defined in IETF RFC 3711, which details the methods of providing confidentiality and data integrity for both RTP voice and video media as well as their corresponding RTCP streams.

In SRTP, encryption is applied only to the payload of the RTP packet using an Advanced Encryption Standard (AES) algorithm with a 128-bit key. SRTP also uses HMAC-SHA1 as the message authentication hash algorithm. Message authentication is applied to the RTP header as well as the RTP payload. SRTP protects against replay attacks by applying the message authentication to the RTP sequence number within the header.

As with SRTP packets, encryption applies only to the payload of the SRTCP packet, when utilized. Message authentication, however, is applied to both the RTCP header and the RTCP payload.

Datagram Transport Layer Security (DTLS) Secure Real-Time Transport Protocol (SRTP)

The Datagram Transport Layer Security (DTLS) protocol is designed to provide authentication, data integrity, and confidentiality for communications between two applications, over a datagram transport protocol such as User Datagram Protocol (UDP). The protocol is defined in IETF RFC 4347. DTLS is based on TLS, and it includes additional mechanisms such as sequence numbers and retransmission capability to compensate for the unreliable nature of UDP. DTLS-SRTP is an extension to DTLS for the negotiation of SRTP keying material within DTLS.

In a Cisco Telepresence solution, the DTLS handshake occurs directly between the TelePresence endpoints. The DTLS-SRTP session is established between the Cisco TelePresence codecs, within the RTP media streams between two endpoints, but not with any associated Cisco Unified IP Phone in the call. In each call, two DTLS-SRTP handshakes occur, one for voice and one for video media, and keys are negotiated for encryption and authentication of both streams.

Digital Certificates

The Cisco Unified Communications System uses X.509 v3 certificates as part of its Public Key Infrastructure (PKI) feature for generating public and private keys used for encrypting and decrypting messages. This PKI implementation generates key pairs that can encrypt messages with the private key that can be decrypted only with the public key that is exchanged between two devices. The private key is kept secure within the device and never exposed. The public key is available as an attribute defined on the X.509 digital certificate. The attributes are established by a Certificate Authority (CA), which digitally signs the certificate. The digital signature itself is a hash of the message, encrypted using the private key of the Certificate Authority. The digital signature of the Certificate Authority can be verified by the recipient using the public key of the Certificate Authority.

A certificate can be either a Manufacturing Installed Certificate (MIC) or a Locally Significant Certificate (LSC). MICs are pre-installed and LSCs are installed by the Cisco Certificate Authority Proxy Function (CAPF) on Cisco Unified Communications Manager (Unified CM). The MIC certificate can provide the credentials used by the endpoints to perform a first-time authentication and enrollment into Cisco Unified CM's security framework. When MICs are used, the Cisco CA and the Cisco Manufacturing CA certificates act as the root certificates.

Note The MIC is also used for establishing Datagram Transport Layer Security (DTLS) sessions between Cisco TelePresence endpoints.

Certificate Authority Proxy Function (CAPF)

The Cisco Certificate Authority Proxy Function (CAPF) is a software service installed as part of Cisco Unified CM. CAPF is not enabled by default and needs to be configured after installation. CAPF issues Locally Significant Certificates (LSCs) for Cisco Unified IP Phones and Cisco TelePresence endpoints. CAPF self-signs certificates under its own authority. However, it can be used as a proxy to request certificates from an external Certificate Authority (CA). It supports the signing of certificates by a third-party certificate authority (CA) using Public-Key Cryptography Standard (PKCS) #10 Certificate Signing Request (CSR).

When using third-party CAs, the CAPF can be signed by the CA, but the phone LSCs are still generated by the CAPF. When self-signed LSCs are used, the CAPF certificate is the root certificate. When an external CA is used, the CAPF acts as the subordinate CA, and the external CA is the root CA.

These certificates are then used to establish secure, authenticated connections for protocols such as SIP signaling over TLS.

The Certificate Trust List (CTL)

The CTL Provider is another software service, installed as part of Cisco Unified CM, that works together with a CTL Client to generate a Certificate Trust List (CTL). The CTL Client is a software plug-in that can be downloaded from the Cisco Unified CM server and run on a separate Windows PC. The Certificate Trust List itself is a predefined list of trusted certificates stored on the Unified CM server and downloaded as a file to the Cisco endpoints when they boot up. The CTL indicates the list of Unified CM servers that the Cisco Unified IP Phones and TelePresence endpoints can trust when they initiate SIP sessions over TLS for call signaling. In order to provide authentication for the CTL itself, a minimum of two separate Cisco Universal Serial Bus (USB) hardware security keys (etokens) are required. These USB keys are not part of the Cisco Unified CM product and must be purchased separately. These security keys are inserted into the PC running the CTL Client plug-in during the CTL generation process.

Configuration File Integrity and Encryption

The configuration files for Cisco TelePresence units and Cisco Unified IP Phones are stored within Cisco Unified CM. These files are downloaded to the endpoints each time they boot up. Configuration files are also automatically downloaded to a Cisco TelePresence device any time a change in configuration is made within Unified CM that would affect the endpoint's configuration. A configuration file download also resets the device.

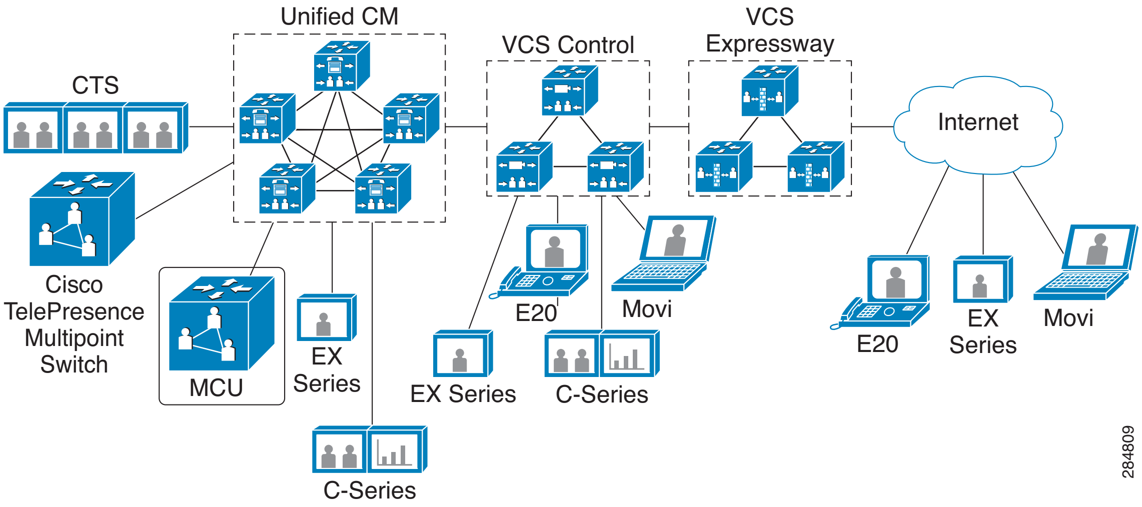
Device security profiles can be created on Cisco Unified CM that require encryption of the configuration files. This prevents configuration files from being changed by unauthorized users because they are digitally signed by Unified CM.

Media Encryption Details

Cisco Unified Communication Manager (Unified CM) supports Secure Real-time Transport Protocol (SRTP) for the audio portion of a voice call payload but does not support encryption for video media. Native support for Cisco TelePresence EX Series and C Series endpoints has been added to Cisco Unified CM 8.6 and later releases, but this does not include support for media encryption. Cisco TelePresence System and Video Communication Servers support SRTP for endpoints natively registered to them. Cisco TelePresence endpoints use Datagram Transport Layer Security (DTLS) for private-key exchange used in establishing SRTP.

Cisco Unified CM, Cisco TelePresence System (CTS), and Cisco Video Communication Servers (VCS) support secure signaling using TLS for SIP. In implementations where a SIP trunk is used for integrating Unified CM, VCS, and CTS, end-to-end signaling encryption of SIP protocol is supported using TLS (see Figure 9-2).

Figure 9-2 Integration of Cisco TelePresence System, Unified CM, and Video Communication Servers Using TLS



Implementing end-to-end SIP signaling encryption requires the configuration of a VCS neighbor zone to Unified CM to use TLS. This feature requires the installation of the appropriate feature key. In addition, Unified CM must be able to trust the VCS server's certificate. This can be done either by having both Unified CM and VCS use certificates from the same Certificate Authority or, if a common root CA is not used, then by exporting the VCS server certificate and uploading it to the Unified CM trust store

While signaling encryption is achieved using this configuration, this does not secure the media payload. Encryption of the video calls requires using DTLS between the endpoints for establishing a secure channel for key exchange. The AES encryption keys that are used for media encryption are then passed through this channel. This media encryption can be implemented on TelePresence endpoints configured to support media encryption but will not work on Unified CM IP Phones.

For step-by-step configuration instructions, refer to the *Cisco TelePresence Video Communication Server Cisco Unified Communications Manager Deployment Guide*, available at

> http://www.cisco.com/en/US/products/ps11337/products_installation_and_configuration_guides_list.html

Integration with Firewalls and Access Control List Considerations

A secure enterprise network relies on firewalls in conjunction with access control lists (ACLs) to protect the network from various sorts of malicious threats. ACLs are also frequently used to enforce Quality of Service (QoS) settings, including marking, shaping, and policing traffic at various places in the network, such as at the access edge of a local area network (LAN) or at the intersection of a LAN and wide area network (WAN). Firewalls may also be used for access control within an enterprise campus and between two or more campus locations.

The servers and endpoints in a Cisco Unified Communications System use a large range of ports and services; therefore, using firewalls and ACLs to protect them and restrict access to them requires careful planning. Given the complexities that firewalls introduce into a network design, care is needed in placing and configuring the firewalls and the devices around the firewalls to allow the traffic that is considered correct to pass while blocking the traffic that needs to be blocked.

Because of the dynamic nature of the ports used by voice and video devices, having a firewall helps to control opening up a large range of ports needed for the different services used by the Cisco Unified Communications System. Application Layer Inspection functionality in firewalls simplifies traffic filtering by dynamically opening and closing required ports and sockets. It performs deep packet inspection to obtain the embedded IP addressing information for establishing media streams in a call. However, for this to function well, the firewall's inspection engine must support the specific protocol implementation of the Unified Communications components. The Cisco Adaptive Security Appliance (ASA) 5500 Series firewalls support version-specific implementations of Unified Communications protocols. This requires that the version of ASA implemented is compatible with the version of the Cisco Unified Communications solution in the network. Upgrading one might require upgrading the other.

The Cisco ASA 5500 firewalls restrict and allow traffic based on the trust levels assigned to its interfaces. This establishes different levels of trust within a network. Security levels range from 100, which is the most secure interface, to 0, which is the least secure interface. These are often referred to as the "inside" and "outside." By default, traffic initiated from a device on an interface with a higher security level is allowed to pass to a device on an interface with a lower security level. Return traffic corresponding to that session is dynamically allowed from the lower interface security level to the interface with the higher security level. This behavior works well with the Cisco TelePresence endpoints that use symmetric port numbering in point-to-point calls. However, multipoint TelePresence calls cannot always use symmetrically numbered ports.

In multipoint TelePresence calls, the audio and video User Datagram Protocol (UDP) streams flow between the Cisco TelePresence endpoints and the Cisco TelePresence Multipoint Switch. The endpoints each have one audio and one video call, but because the Multipoint Switch has a single IP address and has to support multiple UDP audio and video streams from multiple endpoints, the flows are not necessarily symmetric from a UDP port numbering perspective. This requires configuring application layer protocol inspection for SIP protocol in order to allow the firewall to open and close the necessary media ports dynamically.

Firewalls do not allow traffic initiated from a device on an interface with a lower security level to pass to a device on an interface with a higher security level. This behavior can be modified with an ingress access control list (ACL) on the lower security interface level. An ingress ACL applied to the interface with the higher security level may also be used to limit traffic going from higher level security interfaces to interfaces with lower security levels.

Cisco ASA 5500 Series firewalls can also be allowed to operate with interfaces having equal security levels. This requires configuring commands that permit same security interface traffic. ACLs can also be applied on each interface, and static translations can be used to specifically allow access between certain devices and protocols connected to interfaces with equal security levels.

For a list of TCP and UDP ports that need to be permitted between Cisco TelePresence components, refer to the *Securing Cisco TelePresence Products* document, available at

http://www.cisco.com/en/US/products/ps7315/products_installation_and_configuration_guides_list.html

For a list of ports used by Cisco Unified CM, refer to the *Cisco Unified Communications Manager TCP and UDP Port Usage* guide, available at

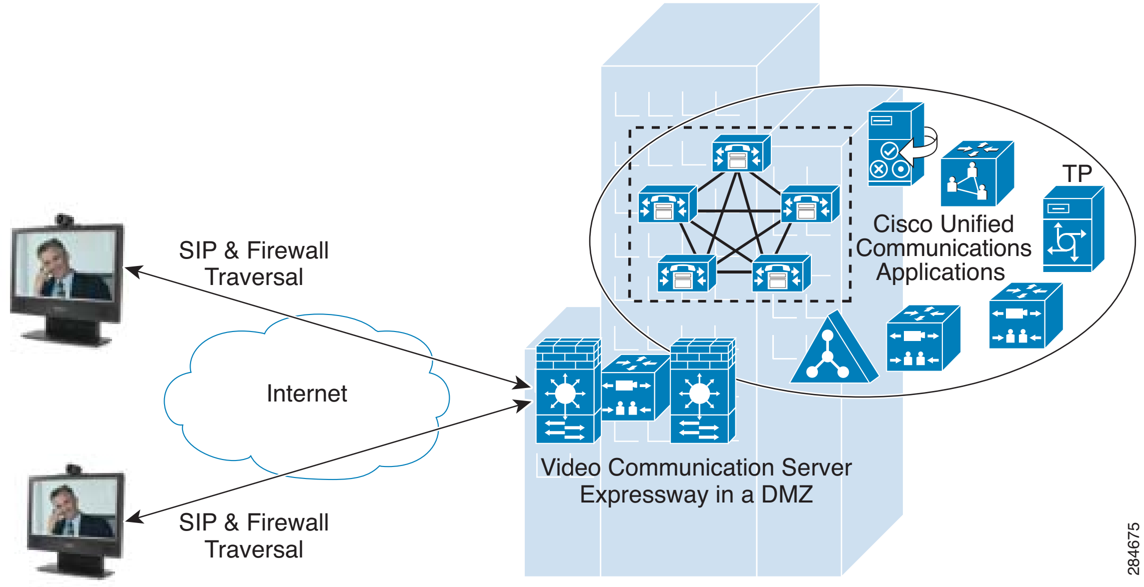
http://www.cisco.com/en/US/products/sw/voicesw/ps556/prod_maintenance_guides_list.html

Firewall Traversal in the DMZ

The Cisco TelePresence Video Communication Server Expressway (VCS Expressway) can establish video communication calls with devices outside the enterprise network and across the Internet. The VCS Expressway must be placed outside the private network used by the Cisco Unified Communications solution to allow external callers to access the device. It can be deployed either on the public Internet or in a demilitarized zone (DMZ). By default, firewalls block unsolicited incoming requests, so the firewall must be configured to allow the VCS Expressway to establish a constant connection with the VCS Control server.

Positioning the VCS Expressway in the DMZ makes this implementation much more secure (see Figure 9-3). It uses VCS as the dedicated server for handling voice and video traffic, thus making the firewall configuration less complex. It can limit the management traffic to the VCS Expressway, restricting it to the internal private traffic and blocking access externally.

Figure 9-3 VCS Expressway in a DMZ


SIP & Firewall Traversal
Internet
SIP & Firewall Traversal
TP
Cisco Unified Communications Applications
Video Communication Server Expressway in a DMZ
284675

GLOSSARY

Revised: March 30, 2012, OL-27011-01

A

AES	Advanced Encryption Standard
ASA	Cisco Adaptive Security Appliance

B

B2B	Business-to-business
BFCP	Binary Floor Control Protocol
bps	Bits per second

C

CA	Certificate Authority
CAPF	Cisco Certificate Authority Proxy Function
CBWFQ	Class-Based Weighted Fair Queuing
CDP	Cisco Discovery Protocol
CIF	Common Intermediate Format
codec	COmpressor-DECompressor (or COder-DECoder)
CSR	Certificate Signing Request
CTL	Certificate Trust List
CTS	Cisco TelePresence System

D

DCT	Display Channel Table
DHCP	Dynamic Host Configuration Protocol

DID	Direct inward dialing
DiffServ	Differentiated Services
DMZ	Demilitarized zone
DNS	Domain Name System
DNS SRV	DNS Service
DSCP	Differentiated Services Code Point
DSP	Digital signal processor
DTLS	Datagram Transport Layer Security

E

ENUM	Electronic Numbering

F

FECC	Far-end camera control
FIFO	first in, first out
fps	Frames per second

G

GDR	Gradual Decoder Refresh

H

HD	High definition
HMAC	Hash-based Message Authentication Code
HTTP	Hyper-Text Transfer Protocol
HTTPS	Hyper-Text Transfer Protocol Secure

I

i	Interlaced video format
IDR	Instant Decoder Refresh
IEC	International Electrotechnical Commission
IETF	Internet Engineering Task Force
IME	Cisco Intercompany Media Engine
IP	Internet Protocol
IPsec	Internet Protocol Security
IPv4	IP version 4
IPv6	IP version 6
ISDN	Integrated Services Digital Network
ISO	International Organization for Standardization
ISR	Cisco Integrated Services Router
ITU	International Telecommunication Union
ITU-T	International Telecommunication Union Telecommunication Standardization Sector

L

LDAP	Lightweight Directory Access Protocol
LSC	Locally Significant Certificate
LTRP	Long-Term Reference Picture

M

MAC	Message Authentication Code
MCU	Multipoint Control Unit
MIC	Manufacturing Installed Certificate
MPEG	Moving Picture Experts Group
MPLS	Multiprotocol Label Switching
MSE	Cisco TelePresence Media Services Engine

MSI	Cisco Media Services Interface
MXE	Cisco Media Experience Engine

N

NAT	Network Address Translation
NAT-T	Network Address Translation Traversal
NTP	Network Time Protocol

O

OBTP	One Button To Push

P

p	Progressive scan
PKCS	Public-Key Cryptography Standard
PKI	Public Key Infrastructure
PQ	Priority Queuing
PVDM3	Cisco High-Density Packet Voice Video Digital Signal Processor Module

Q

QCIF	Quarter Common Intermediate Format
QoS	Quality of Service

R

RFC	Request for Comments
RSVP	Resource Reservation Protocol
RTCP	Real-time Transport Control Protocol
RTP	Real-time Transport Protocol

S

SCCP	Skinny Client Control Protocol
SD	Standard definition
SIP	Session Initiation Protocol
SME	Cisco Unified Communications Manager Session Management Edition
SRTCP	Secure Real-time Transport Control Protocol
SRTP	Secure Real-Time Transport Protocol
SSH	Secure Shell
SSL	Secure Sockets Layer

T

TCP	Transmission Control Protocol
TCS	Cisco TelePresence Content Server
TLS	Transport Layer Security
TMS	Cisco TelePresence Management Suite
ToS	Type of Service
TURN	Traversal Using Relay NAT

U

UAC	User agent client
UAS	User agent server
UDP	User Datagram Protocol
Unified CM	Cisco Unified Communications Manager
Unified CST	Cisco Unified Communications Sizing Tool
URI	Uniform resource identifier
USB	Universal Serial Bus
UTF	Unicode Transformation Format

V

VCL	Virtual Channel Link
VCS	Cisco TelePresence Video Communication Server
VLAN	Virtual Local Area Network
VoIP	Voice over Internet Protocol

W

WFQ	Weighted Fair Queuing

INDEX