C H A P T E R **3**

Basic Video Concepts

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:

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
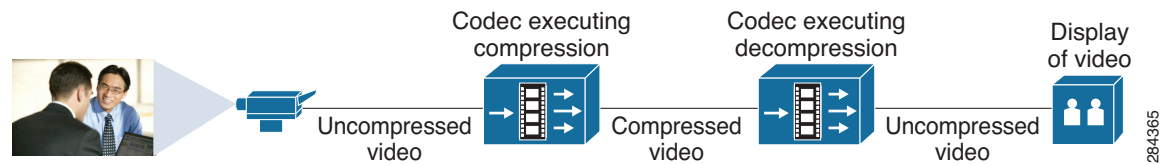
Video Compression Formats

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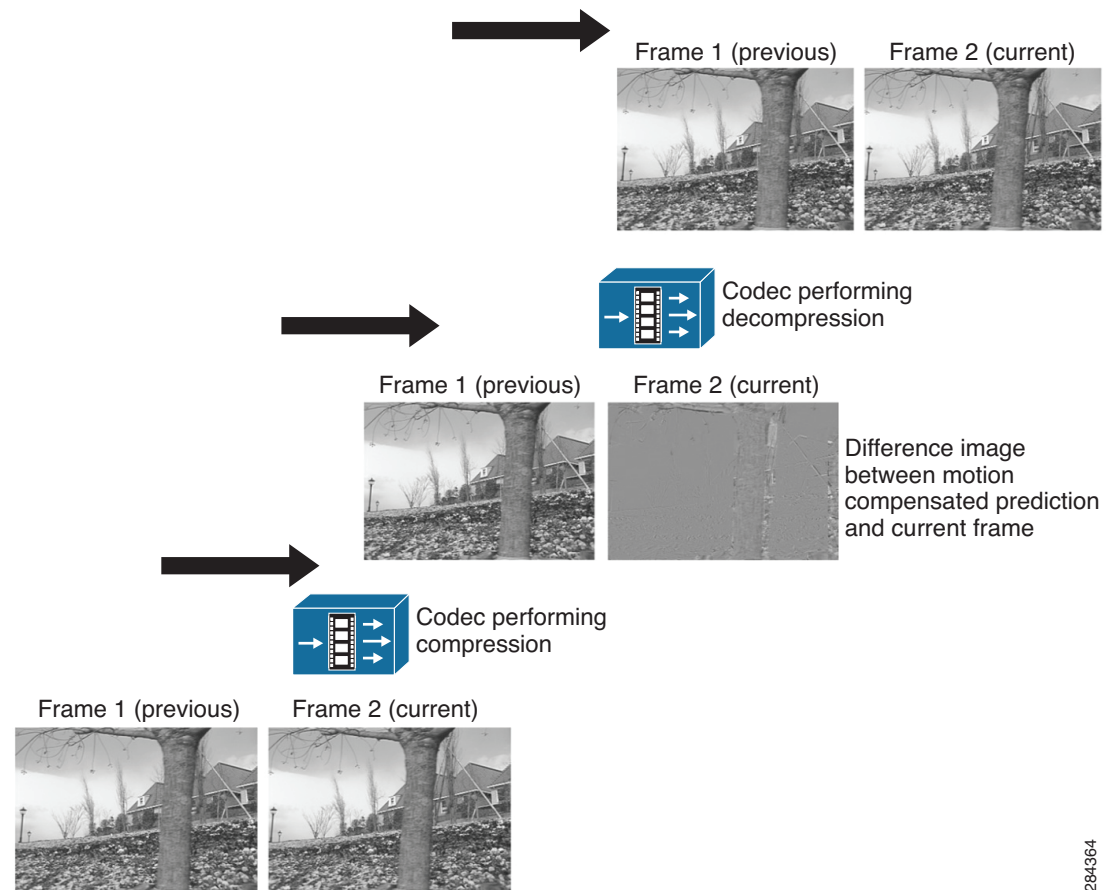
There are many different kinds of compressed video frames used in IP video solutions, but the main types are:

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.
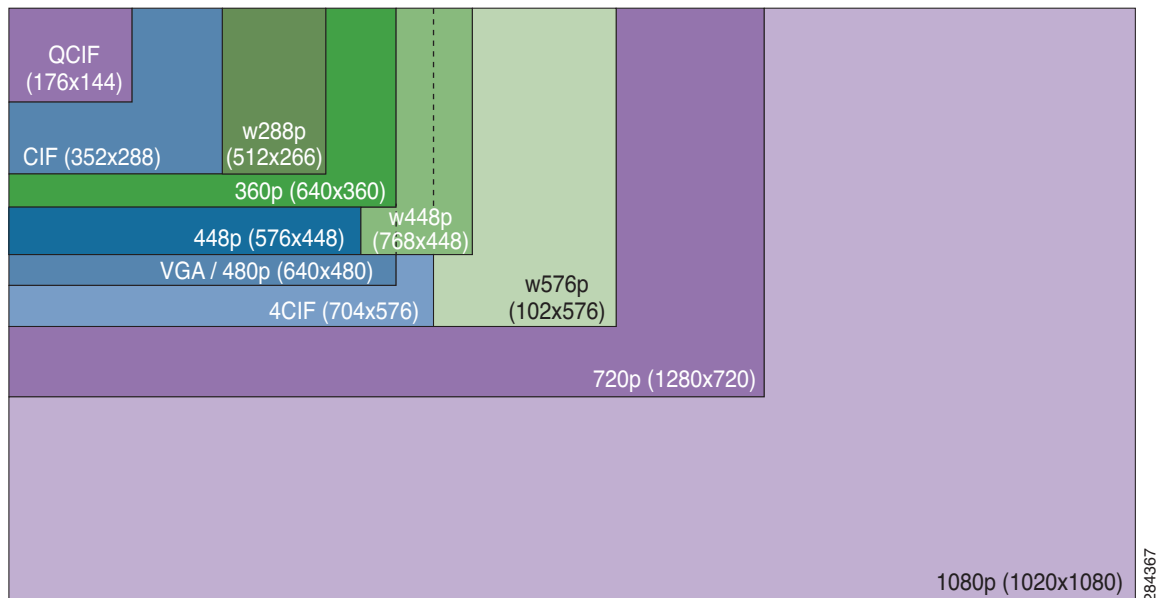
Figure 3-3   Order of Video Display
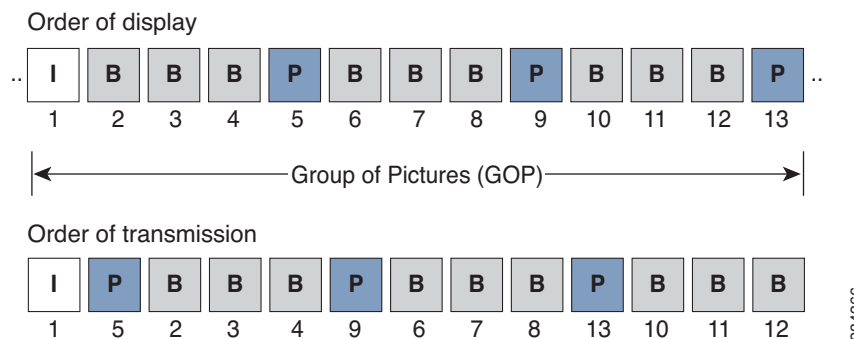
Resolution Format in IP Video Solutions

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

QCIF (176x144)
CIF (352x288)
w288p (512x266)
360p (640x360)
w448p (768x448)
448p (576x448)
VGA / 480p (640x480)
w576p (102x576)
4CIF (704x576)
720p (1280x720)
1080p (1020x1080)
284367

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:

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.
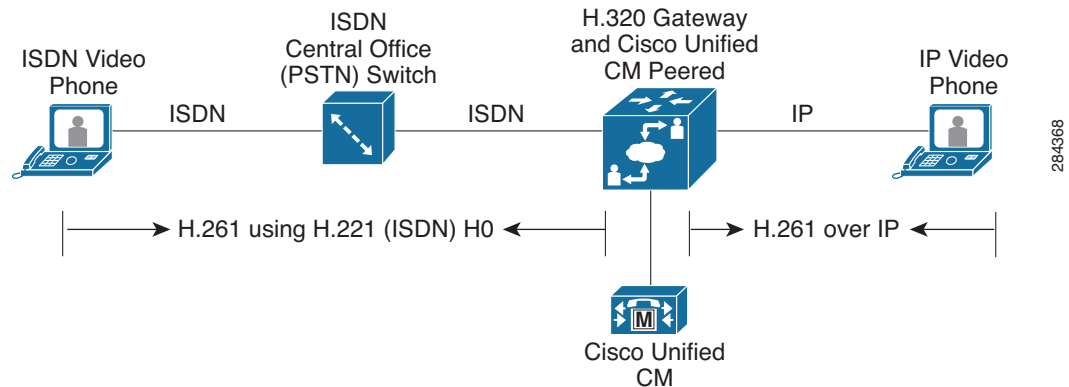
Video over ISDN

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*

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*
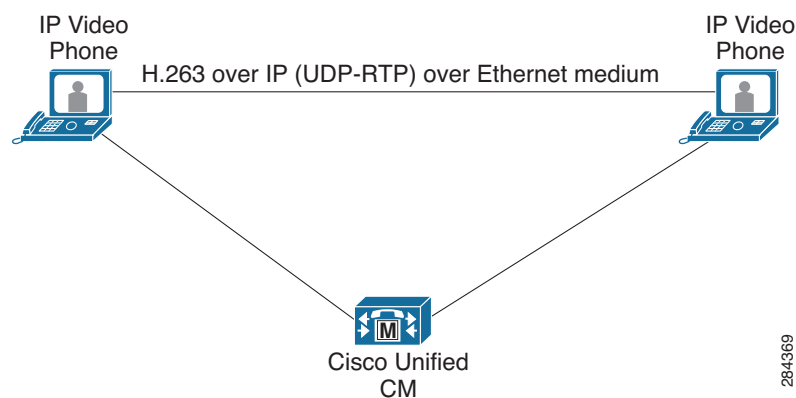
As shown in Figure 3-7, IP Video Telephony improves the video resolution by providing transmission flexibility in a variety of physical transport media that are not restricted to 2 Mbps as video over ISDN was (1.54 Mbps in the case of the US). IP (User Datagram Protocol for video) encapsulated packets 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

| Ethernet	IP (UDP-RTP)	H.264 – Video Payload
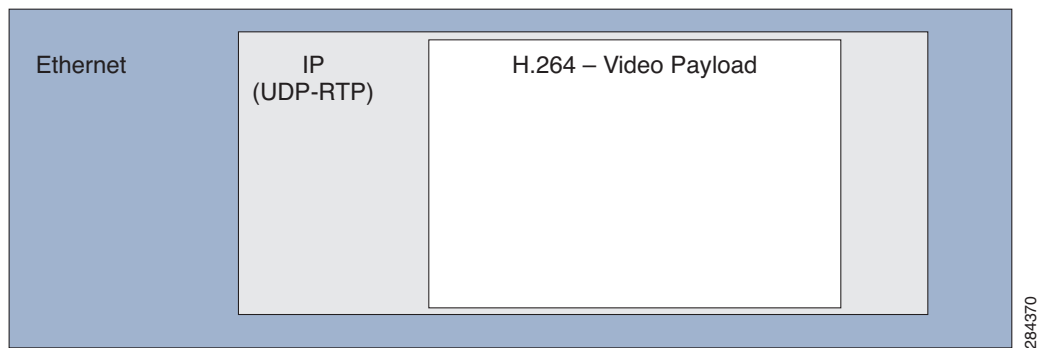
Desktop Video Conferencing

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*

Network/Cloud
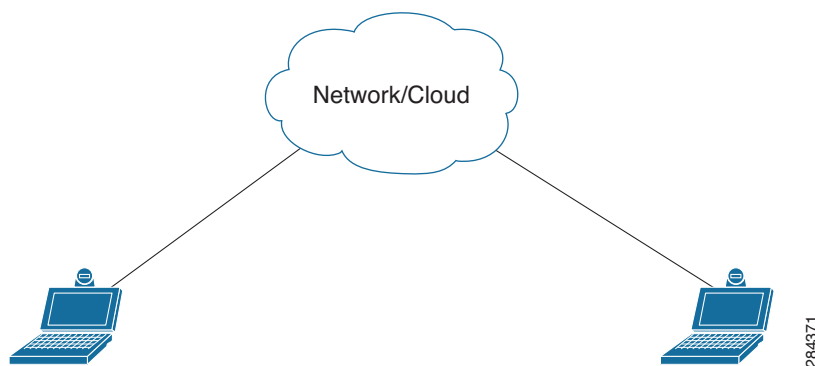
Immersive Video Conferencing

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
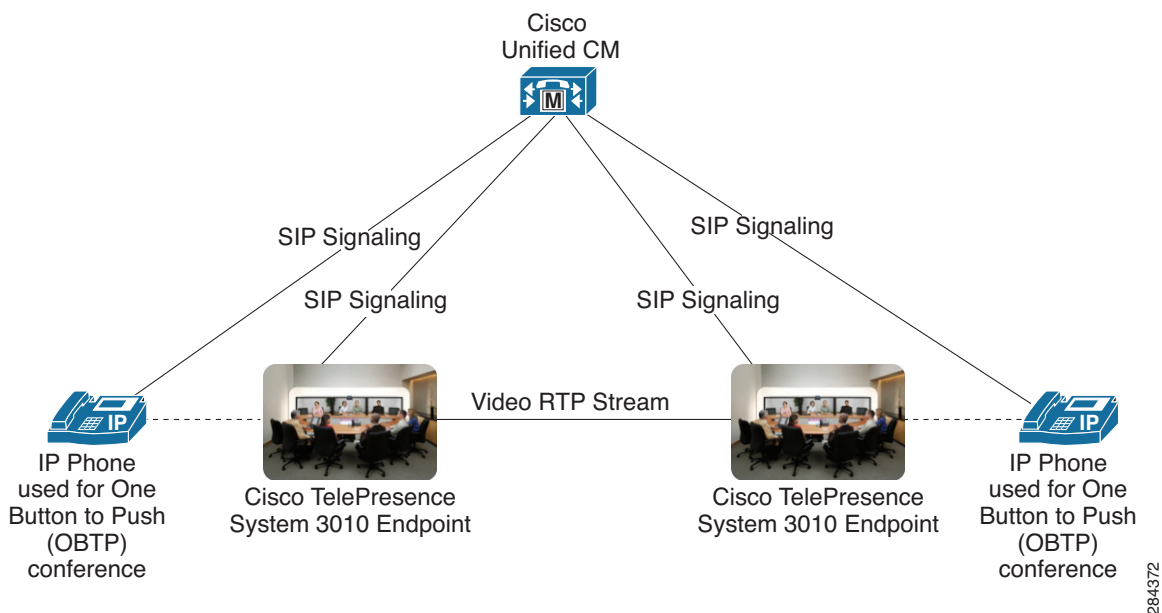
Cloud-Hosted Video Solutions

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
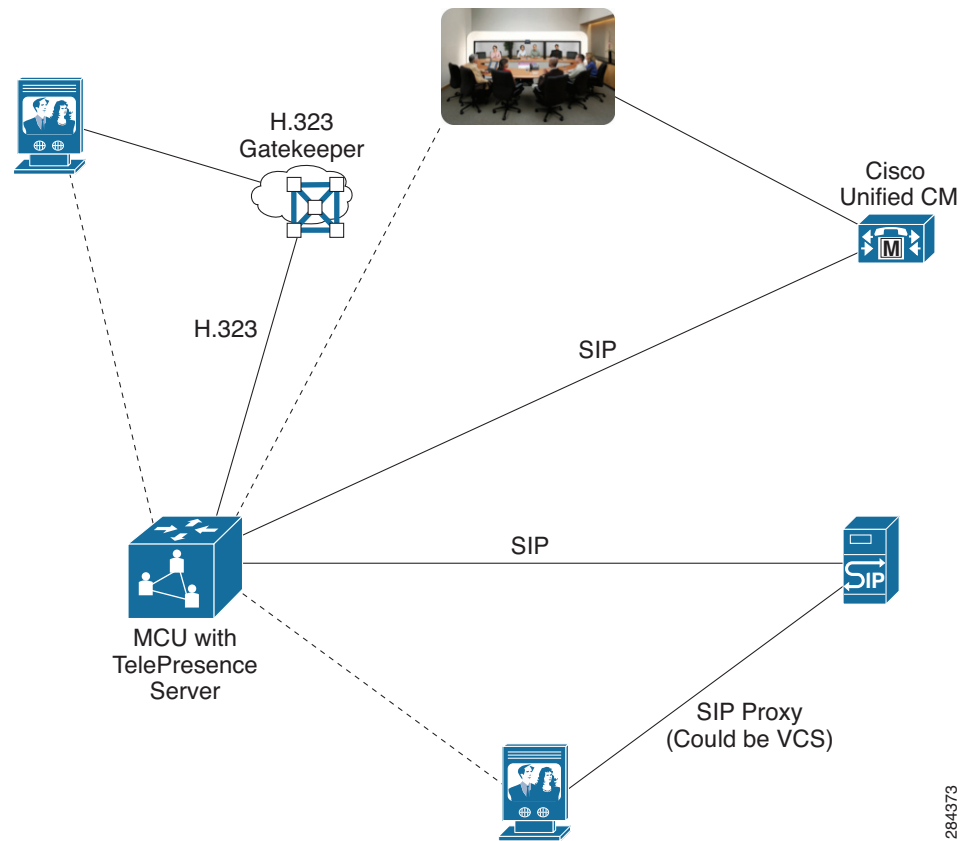
Legacy Multipoint Control Units

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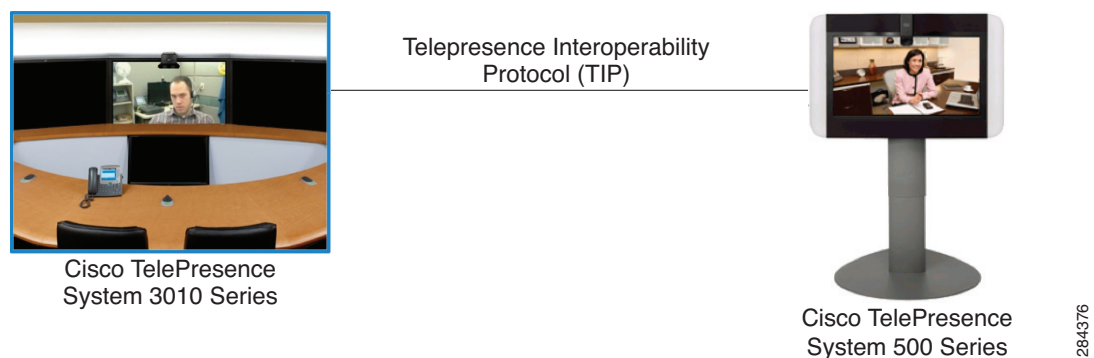
- Telepresence Interoperability Protocol (TIP), page 3-14
- ClearPath, page 3-15

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

Cisco TelePresence
System 3010 Series

Telepresence Interoperability
Protocol (TIP)

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.