Interactive Video on Enterprise Networks

By Christine Perey
PEREY Research & Consulting

www.perey.com

This document summarizes issues and solutions generally available to the public as of October 16, 2000.

The information herein is provided by PEREY Research & Consulting as a service to the industry and, with proper attribution to its original source, may be cited and republished in its entirety, electronically and in print.

Neither the author nor the publisher shall be liable for any errors of fact or judgment or for any damages resulting from reliance upon this information.

Table of Contents

Introduction

Network convergence is a popular concept. It conveys the merging of network protocols and media (data, voice and video) into a single, centrally managed intelligent unit. This system will reach from palm and desktop to conference room and theater, regardless of the physical distances between these sites, to offer data as well as rich media content, creating compelling user experiences.

Network architects recognize that multimedia over IP implementation details will vary from site to site, depending on a combination of internal business requirements and the unique conditions in a network at the time multimedia communications support goes in. This document is designed to assist those who seek to introduce interactive video to their corporate IP network users as a first step towards network convergence. On the one hand, it is sufficiently general to apply to a variety of networks. On the other, it contains details on quality of service (QoS) protocols, products and other elements of network architecture that have not heretofore been compiled in a single resource. The network manager may use this document when outsourcing certain tasks such as network audits, design and implementation of networked multimedia.

Framework for Implementation

Based on years of dialog with the manufacturers of multimedia network components and first-hand experiences of those who have deployed voice and video in large and small networks, this white paper provides a framework for an enterprise-wide video-over–IP (VoIP) deployment process.

This framework approaches VoIP implementation in the following five stages:

1. Analyze needs and locate the most multimedia-hungry applications in the enterprise, establish which type of video is needed and where and then develop a deployment plan accordingly.
2. Examine and address weaknesses in all or selected network segments identified by the needs analysis to do the following:
   - Rework topology and accommodate the increased bandwidth for business-quality video to selected locations
   - Provide QoS guarantees to interactive audio and video traffic without shortchanging data-only applications
   - Monitor and control network utilization by application
3. Select the endpoints and servers that best match the application requirements and the network.
4. Address user-to-user and business-to-business connectivity issues.
5. Refine the total solution to the evolving needs of users and customers.

1. Needs Analysis

Before investing in products and services for video over IP in an enterprise, an IT executive and representatives of different parts of the business, possibly including outside contractors, should form a team to assess multimedia networking needs and resources. This group should review corporate communications objectives as well as other relevant business goals (e.g., moving to e-business practices, retaining valuable knowledge workers and increasing productivity). A needs-analysis phase should also survey one or more user groups about how they perform their functions today and their willingness to change behaviors to maximize the impact of new technologies on their productivity.

In this process, it will be important to distinguish clearly the level of interactivity users expect to have with their multimedia content. Users will have applications that warrant the following:

- control over one-way (broadcast or on-demand) video streaming from a server
- an asymmetrical bit rate but highly responsive level of speaker/audience interaction
- two-way interactive video (videoconferencing)

Furthermore, an analysis will differentiate between applications that enhance communications:

- within one facility
- between remote offices of the same enterprise
- with partners, suppliers and customers (B2B and B2C)

In the analytical phase, the business process managers will also identify the applications that will best leverage a common voice or video service in the local (LAN) or wide area networks (WANs). A “service” in the present paper is defined as a network-based, managed capability that provides value to the day-to-day operations of a group of people. The service can be embedded in the infrastructure (routers and switches) or ride above. This paper is about the implementation of interactive (2-way) video-rich service and is not intended to cover all aspects of service provisioning/delivery through service-specific environments or facilities.

While the focus of the present paper is on *video-enabled* applications, we assume that the voice infrastructure in place in businesses today (another two way interactive service) is fulfilling many mission-critical communications functions and will likewise receive the attention of a multidisciplinary network convergence strategy team.

Return on Investment

It is important for the multidisciplinary team to identify the ways in which the introduction of technology will impact the business and how this impact will be measured. For example, cost savings on travel is the first place most companies target. High value employee and client/customer retention, especially among professional service providers, is another common business metric that benefits as a result of expanding visual communications in the enterprise.

Typically a return on investment (ROI) calculation begins with the measurement of basic and advanced activities or processes in place prior to the introduction of new technologies. Once the measurement has been done on the baseline activities and post-deployment activities, a project may establish with greater credibility that it has increased efficiencies or had a positive ROI. At such a point, the project is likely to garner additional resources.

A few large organizations already enjoy significant cost savings from having moved voice to a WAN for communications between remote office telephone systems (avoiding tariffs to local and long-distance circuit-switched service providers). These private branch exchange (PBX)–to–PBX trunks do not commonly use IP as the transport.

Voice/data convergence on the LAN is not widely implemented due to the number of changes required and the high risk of disrupting a mission-critical service such as telephony[1]. When VoIP on the LAN is as reliable as today's PBXs, it will take hold. In the interim, introducing voice and video in selected applications to a restricted number of locations on a LAN does not jeopardize existing processes and offers strategic advantages.

At the end of the needs-analysis phase, the team will communicate their findings and recommendations, including measurable objectives and a target list of sites for the first-phase implementation, to a deployment team. Technical deployment team members will have audited the enterprise network for multimedia applications. They will weigh the capabilities of products and services available (at the time deployment is to begin) against the results of the audit, corporate goals and what vendors have promised. Then an evolutionary deployment strategy will follow.

The Interactivity Spectrum's Impact on Network Design

The deployment plan will reflect the level of interactivity the applications will require. Most professionals recognize that communications objectives such as corporate branding, advertising, investor relations, learning and other professional development services are more compelling and accessible for busy audiences on a PC screen than when printed in a brochure or manual. This is primarily because, with a few mouse clicks, anyone can interact with and customize the information they receive on a screen. And when presented on the screen, information impact is greatly enhanced when accompanied by video and audio content. Visual and auditory communications can transfer ideas more quickly and fully and leave a longer-lasting impression on the viewer.

The challenge for communications professionals is to select the optimum level of interaction between the audience or viewer and the content source. It would be convenient to be able to pinpoint two or three levels of interactivity, but, in fact, a continuum more fully represents the opportunities and the challenges.

Television (or business television via satellite) broadcast defines the left end of the continuum in which once a channel is selected, the user passively watches information. Moving to the right, one enters the streaming-media realm. When a previously recorded and compressed file is retrieved from a media library, the user expects a slightly greater, yet limited level of interactivity. For example, streaming-media interfaces typically permit the user to begin file delivery (play), pause and to navigate within a file (by manipulating a marker on the progress bar). But, during the session the viewer can only

[1] According to a Sage Research study published in NetworkWorld May 8, 2000, only 6% of 231 large IT organizations say that they have implemented any VoIP. Inhibitors included lack of maturity of the technology, concerns about interoperability and the potentially high costs of products.

passively watch. Previously recorded presenters cannot receive questions and reply immediately or adapt content in response to issues or opinions offered by the audience.

A streaming-media broadcast (also known as a Webcast or conference cast) permits this intermediate level of interactivity. In this case, the media is streamed at a designated time to an audience convened in cyberspace. The hosting network may offer viewers use of its telephone, e-mail or an interface in the browser to interact with the speaker(s). This offers significantly higher interactivity than an on-demand or playback-only scenario, but the user is unable to stop, replay or “search” the content during the session. In the appendix entitled Streaming Media, the reader will find network-design guidelines for implementing one-way video. In general, tolerance for variable bandwidth and data loss is higher with on-demand video applications than with intermediate levels of interactivity.

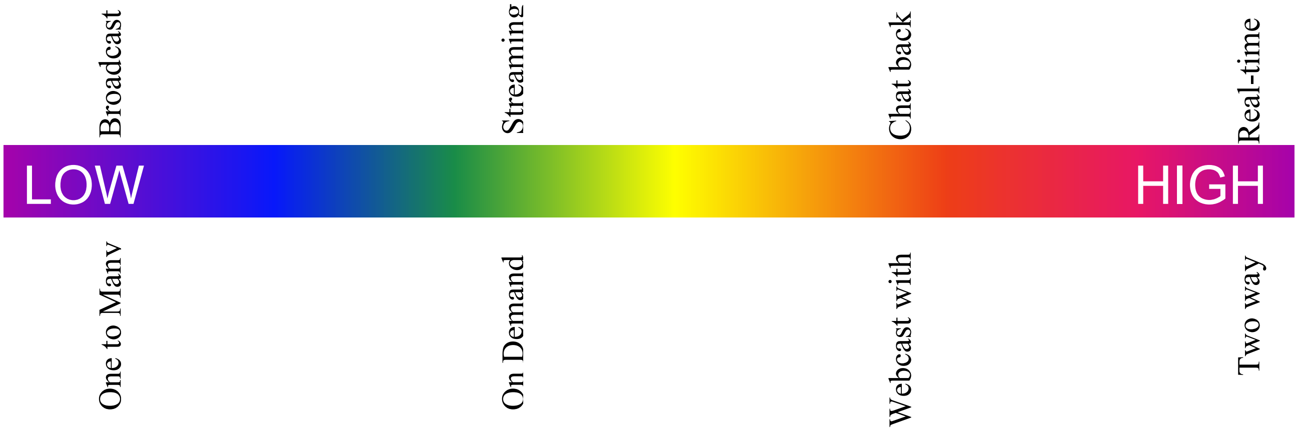


Figure 1. Continuum of Application Interactivity in Visual Communications

In the case of management reviews, collaborative design meetings, sales calls and other forms of communication in which participants are both content sources and viewers, interactivity must be at its highest. Anything less than full interaction compromises the very objective of the session.

Increasing the impact (clarity, size and frame rate) and speed of interactivity adds complexity to network design. Two-way or “fully” interactive video, otherwise known as videoconferencing, requires services that most data networks were not designed to offer.

2. Fully Interactive Video Network Requirements

When multimedia file is streamed to the viewer, as described above, the network infrastructure needs are less stringent than when the audio or video transmission is two-way or “conversational.” For example, when users expect a nearly perfect simulation of the interaction they would have if sharing the same room, there can be no more than 400 ms perceived delay between the time when a person speaks and the moment that it is heard. Video and the associated audio must be captured, compressed, decompressed and synchronized at the receiving terminal at the rate of 30 frames per second.

Capturing, compressing, decompressing and displaying video and audio are the activities performed by endpoints. However, there are many other components to a complete interactive video network. These are integrated with the transport layer in the network. End-to-end network transmission causes packets to be lost and jitter and delay to be introduced. The degree of jitter and packet loss is also influenced by the endpoints’ behavior.

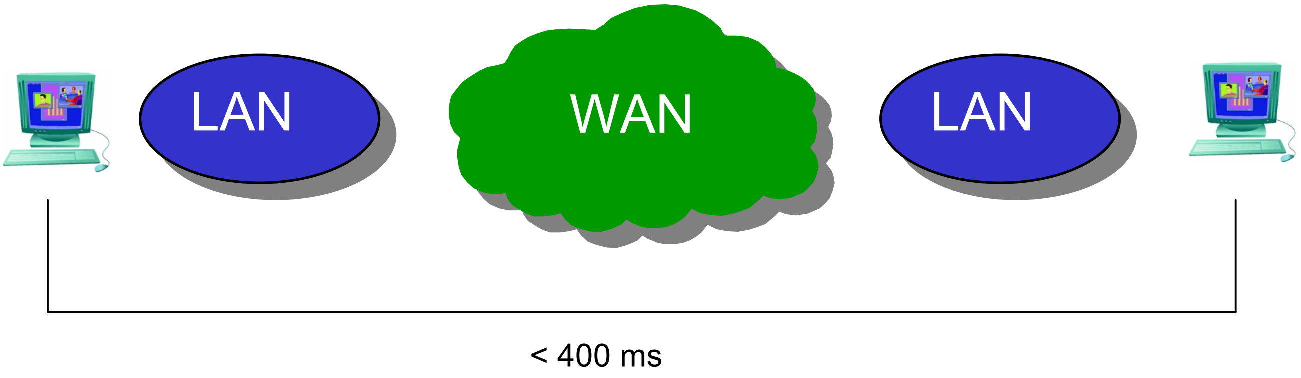


Figure 2. End-to-end network components impacting perceived delay

The trick to optimizing a data network for interactive video over IP is to modify the transmission infrastructure so that during a call there is never more than 400 ms end-to-end delay. This requires management of several variables impacting delay: jitter, network delay, and endpoint performance.

Beginning with the content capture in the near-end device, the system-level view will describe how different components will affect the perceived delay when the audio and video appear on the far end-user display.

Delay is the absolute time for a single IP packet to travel from source to destination. For interactive video applications, this is but one of the components contributing to the perceived video delay.

Jitter is the variation in the delay introduced by network congestion and processing in the network during the transmission of data.

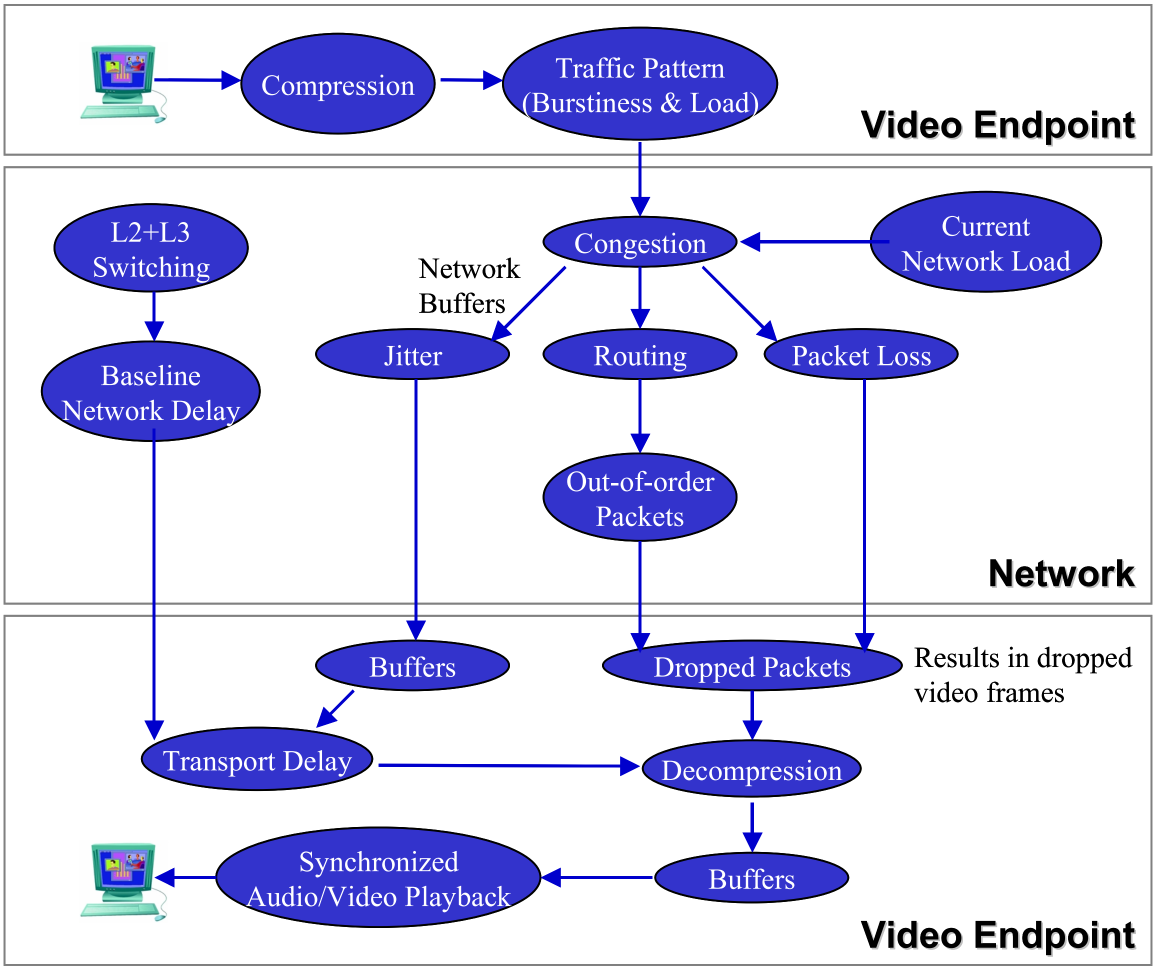


Figure 3. Components and variables contributing to perceived delay **Source: FVC.COM**

Once compressed, audio and video are in independent data streams. These streams meet the network interface where the information is packetized and sent to the network. Here the variables are burstiness and load. Burstiness refers to how much the endpoint controls or shapes the packets to control the time interval between the packets leaving the endpoint and entering the network. Load is the amount of packets put on the network in one second (bandwidth).

Now on the production network, layer 2 and layer 3 switching latency introduces a baseline network delay. In addition, the audio and video packets may experience congestion if, even for a microsecond, when combined with the current network load, the total exceeds the bandwidth. Congestion is handled in routers and switches by buffers. The time spent in the buffer and the rate at which the packets leave the buffer differ. This results in jitter.

When the network components exceed the capacity of buffers, packets are lost. In addition, congestion affects the route the packets may take through the network. This may result in packets arriving at the endpoint interface out of order (sequence at arrival differs from their departure sequence). Most endpoints drop out-of-order audio and video packets.

The total transport delay, once in the endpoint, is the sum of the delay from switching and the delay introduced by buffers in the endpoint to accommodate network jitter. At this point, packets are returned to audio and video data streams that can be decompressed by the respective algorithms (codec).

Decompression in the endpoint produces audio and video data. The time necessary to decompress the audio and video will vary, and buffers in the application ensure that the audio and video data are synchronized for playback.

Reducing network congestion is the key goal to reducing network delay and jitter. Ideally, a network engineer will use a three-pronged approach to reducing congestion and ensuring predictable, but low delay necessary for business-quality interactive video. These prongs include the following:

- Reworking network topology to provide ample bandwidth
- Implementing QoS standards on the network and making sure the WAN carriers do so as well
- Multicast versus unicast
- Taking measures—with appropriate video-enabling network components—to monitor network utilization and control caller access to services, if and when congestion occurs

Reworking Topology to Guarantee Bandwidth

Most companies planning to put video and audio over IP begin by deploying 10/100 switched Ethernet throughout their LANs. The multimedia applications will not require all this bandwidth all of the time, but this architectural change to switched networks is linked to a number of other topological choices outlined below. Before discussing the placement of switches relative to user populations, it is important to understand the raw bandwidth requirements for interactive video.

The bandwidth "sweet spot" for interactive video communications is in the 300K to 400K bit/sec per stream range[2]. Conventionally, the bit rate includes audio and video (media) data as well as control signaling. The H.323 protocol, an International Telecommunications Union (ITU) standard for voice or videoconferencing over IP, does not require that two or more endpoints in a session send the same data rate they receive. A low-powered endpoint may only be able to encode at a rate of 100kbps, but, because decoding is less processor-intensive, it could decode a 300kbps video stream.

[2] In full-duplex networks such as ISDN, Ethernet, ATM, and time division multiplexed networks, capacity is expressed as bandwidth in one direction, though an equal bandwidth is available for traffic in the opposite direction. Duplex networks are not necessarily limited to symmetrical bandwidth usage. Sometimes videoconferencing bandwidth usage is expressed as the sum of the bandwidth in both directions. This causes confusion in the IT industry. If bandwidth utilization is symmetrical (e.g., 384K in both directions), the correct way to express this is as a 384kbps session. If the videoconference is asymmetrical (e.g., 384k in one direction and 128k in the opposite), the conference bandwidth utilization should be expressed using a convention such as 384/128kbps for the above example.

Overhead for IP packet headers must also be taken into consideration. For example, a bi-directional 384kbps videoconference will consume approximately 425kbps in each direction of bandwidth on a LAN. A T-1 offers 1.5Mbps in each direction and would be ample bandwidth for two or three 384kbps videoconferences, depending on the amount of simultaneous traffic on the network.

With regard to topology, the objective is to reduce the perceived end-to-end delay and odds of congestion. There is debate about the best topology to accomplish this. Some recommend that the network design minimize the risk of video and data contention by aggregating video endpoints onto common network segments.

Under this design, data-only endpoints are removed from video-only switches and connected to Ethernet switches with non-video endpoints. This involves building parallel logical video and data networks, and, although that can be costly, it will eliminate all risk that video and data will battle for bandwidth in the local loop. Minimal IP readdressing may be required if Dynamic Host Configuration Protocol or BOOTP is not used to assign IP addresses. These two networks are also capable of communicating to provide data connectivity throughout.

To handle real-time applications, video-only switches should be non-blocking and may need high-capacity back planes for minimum buffers and latency. Precautions at the level of a video network gatekeeper may be put in place to prevent too many video calls from going through a switch at the same time. Rather than causing congestion on a network segment, the caller whose session would have exceeded the switch throughput will encounter the equivalent of a busy signal.

Alternatively, video endpoints and data-only endpoints can be mixed on a given switch to reduce the possibility of having too many simultaneous video sessions passing through a given switch, avoiding potential congestion. The logic is that common data applications such as Internet browsing and e-mail only produce intermittent "well-behaved" TCP/IP traffic that can mix without interfering with the continuous UDP traffic generated by the video applications.

When network congestion begins and packets are being dropped, the data applications begin to resend lost packets. As packet loss increases, TCP/IP applications increase the interval between packets, reducing their relative impact on the interactive video session. The greatest risk of this deployment is that as traffic congestion at the switch increases, mission-critical TCP/IP applications back off so far that they time out and eventually terminate.

Network-intensive applications such as manipulating large databases or drawings (for example, CAD/CAM) are more tenuous. Streaming video servers also use UDP for transmission and do not back off when congestion arises because they do not detect packet loss. Streaming applications, therefore, tend to have a high impact on interactive video application quality if they both occur on common network segments.

The most common way to manage the mixture of data and video applications on a converged network is to over provision. This actually means provisioning to meet the load at peak usage times. While the solution is expensive over Internetworking segments and the WAN, it is a very cost-effective way to avoid bandwidth shortages in the LAN.

Also, network designers must consider the impact of having different local-loop (last-mile) connectivity at different sites. ADSL at 384/128kbps may be important to some users. Most H.323 systems tune symmetrically to the bandwidth that is available on the network. Asymmetrical traffic management is an emerging feature in videoconferencing since the rise in popularity of IP as a signaling protocol. This feature will be important to users that have a high bandwidth connection at a central site where they are distributing information to remote sites. The remote sites will receive very good video from the main site or possibly reasonable video from several remote sites.

Whether symmetrical or asymmetrical, it important to keep congestion off of internetworking routers. Using a network analysis tool can nip problems in the bud. A network administrator can initiate sessions and see the traffic in all segments between the sender and receiver endpoints. The application quickly identifies where bandwidth is limiting, where packets are being dropped or where buffers are not adjusted correctly to manage real-time data needs. Engineers can then take necessary steps to upgrade software or increase memory as needed.

But when there is congestion due to peak usage or bursty traffic that can't be addressed by increasing the raw bandwidth, the network manager will certainly want to implement QoS and traffic-shaping technologies.

QoS Provisions: Locally and Wide Area

Quality of service has received considerable attention in data network journals for years, yet it remains poorly understood and, due to lack of consensus in the industry about which of several strategies is superior, many network managers have hesitated to implement one QoS measure over another.

RSVP, the Reservation Protocol, is one of the mechanisms available for QoS on most routers on the market today. However, two more popular prioritization schemes commonly discussed with the topic of QoS are Diff-Serv and IP Precedence.

The bottom line is that implementing QoS in a LAN helps to protect the integrity of service-sensitive applications and does not require forklift upgrades. Most of the leading network equipment vendors already support common QoS standards, such as RSVP; they only need to be enabled by the network administrator.

There is one caveat, however. If the protocol or scheme chosen for QoS guarantee in the local loop is not the same as that implemented in the backbone, the enterprise network needs to put QoS translation software in place for QoS requests to operate end-to-end during a videoconference.

Even when QoS protocols are in place, more is needed for interactive video applications to take advantage of the mechanisms without detrimental effect to mission-critical data applications. By simply prioritizing video over data, the data application performance risks being sacrificed. To avoid this, a network manager should segment and manage bandwidth on each switch and router to limit the total prioritized video traffic.

For the QoS provisioning to occur on video and audio data specifically, the packets associated with the videoconferencing session must be identified. Identification and classification of packets sensitive to QoS can be done either by the sender (i.e., the software application in the endpoint will do this) or by an application-specific intelligent agent in the network (a policy server will allow the network manager to differentiate between users with the same application). This identification of packets and users can be on the basis of IP address, video number, time of day, or other criteria.

Policy server–based prioritization more closely resembles the philosophy of H.323 gatekeepers (see below). Few enterprise networks have implemented the policy server–based approach internally. However, when service providers reach the point at which traffic and services will be charged according to the class of service provided to a specific address, this could become a more attractive solution for IT managers.

Diff-Serv and IP Precedence

Differentiated Services (Diff-Serv) and IP Precedence are two IETF protocols that enable QoS from within a router or a switch by reading information contained in the type-of-service (ToS) byte of the packet header.

IP Precedence uses priority values to enable the switches and routers to sort packets based on this priority. Eight different priority values can be set on a particular application in the endpoint.

Diff-Serv, on the other hand, assigns QoS classifications to traffic from different applications based on service-level agreements between users and service providers. Currently, two service levels are defined—assured and premium (expedited). Because Diff-Serv aggregates flows into these two categories, it is considered by many to be more scalable than Resource Reservation Protocol, which secures QoS on a per-flow basis.

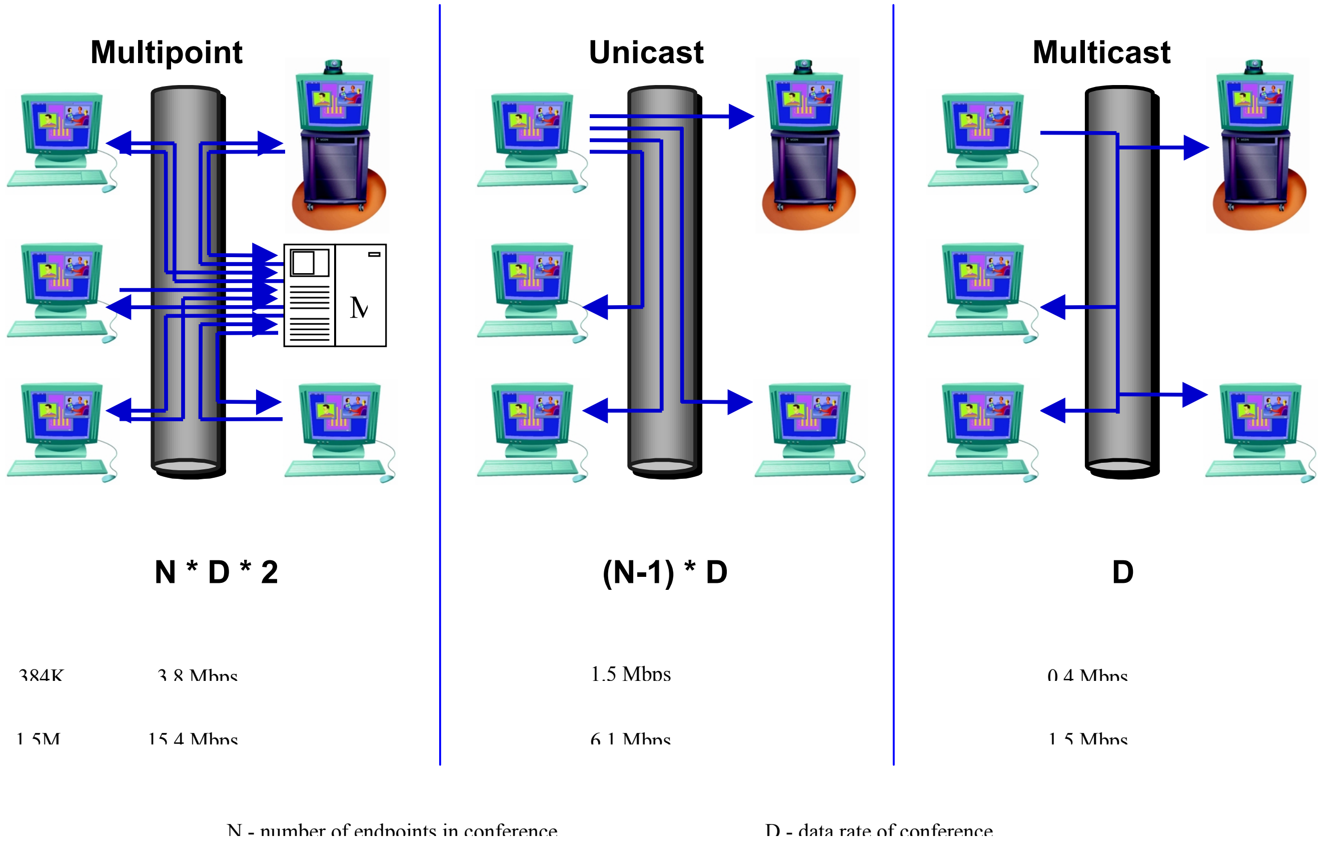
When available, IP Precedence and Diff-Serv will not need to be present in all end-to-end routers to benefit video communication applications. Any routers in your network that support IP Precedence will prioritize packets whose value is set by the endpoint for highest priority in the packet header.

Routers that aren't configured for IP Precedence will give best-effort service to all packets. Network gear vendors are expected to have these protocols implemented fully in the next 12 to 24 months.

In addition to implementing QoS intelligent features in applications and routers in the network, administrators will have to get that same support from the carriers that provide access to WAN links. Although major service providers have long been promising QoS support in their networks, few are using the necessary router-based protocols today.

Multicast versus Unicast

Interactive video packets are usually distributed in unicast mode, from one machine to another or to a multipoint conferencing unit in the network that then unicasts to other machines in a conference. In circumstances where interactive IP video applications are destined for a group of multicast-enabled endpoints with voice-activated switching requirements, the network designer can take advantage of multicast[3] to reduce bandwidth consumption as well as lower CPU load on the sender.



Source: VCON, Inc.

Figure 4. Bandwidth consumption associated with multipoint, unicast and multicast.

When the video-enabled endpoints need send only one packet to an IP multicast group and all participating machines receive the packet, bandwidth consumption is lower than

[3] The range of IP addresses 224.0.0.0 through 239.255.255.255 is meant exclusively for IP Multicast. In the multicast mode, a packet is sent to any machine that wishes to receive it. The sender specifies a multicast address as the destination address in the IP Header. Machines that would like to receive this IP multicast will simply *join* the multicast. The routers in the network will duplicate specific IP packets where and when needed. IP multicast operates at the IP level, and so is best-effort by design.

when an endpoint or multipoint conferencing unit (MCU) sends out multiple copies of the same packet to each of the receivers. Therefore, multicast results in lower network bandwidth usage.

There are, however, many situations in which multicast is not appropriate or introduces an unacceptable level of risk for a two-way interactive video session.

For example, by definition, multicast does not work across non-multicast-enabled routers. For multicast to work over the Internet, for example, all routers would have to forward multicast traffic and propagate multicast routes. This is not a problem for unicast. An IP tunneling application, such as the MBone, can be used to create a multicast-unicast overlay network between subnets and large divided networks such as the Internet.

Multicast architectures cannot accommodate user populations with different bandwidth availability or QoS needs. If signals from one stream receiver to the sender tell the sender to back off, the same reduction in data transmission will be propagated to all endpoints in the meeting. The same situation means that users will not have the ability to fast forward and rewind or seek through multimedia files distributed via multicast.

There are security considerations with multicasting. When a network segment receives a multicast stream in most implementations, by design, every machine on that network can receive the stream.

Monitoring and Controlling Network Resource Utilization

Once the network is designed for a population of video application, the endpoints need to be monitored and their access to pooled resources controlled. In H.323, the gatekeeper is the standard mechanism that provides control over H.323 entities (endpoints, gateways and multipoint control units).

A gatekeeper is software that ensures the smooth operation of an interactive video network. This section explains the principal gatekeeper functions and deployment considerations important to network administrators.

A zone's gatekeeper is logically separate from network endpoints. However, the gatekeeper application may run within any terminal/MCU/ gateway endpoint, or even in a non-H.323 network device such as an NT server. When present in an H.323 network there are three mandatory zone management functions that a gatekeeper **must** perform:

Address Translation—provide address translation between alias and transport addresses upon an endpoint's request for services

Admissions Control—authorize network access based on some specified criteria

Bandwidth Control—monitor and control network bandwidth usage and ensure that the audio and/or video traffic does not exceed maximum network load as defined by the network manager

Regardless of the physical location of the gatekeeper program code, **there must only be one active *runtime gatekeeper* per zone.** The choice of gatekeeper placement is critical to the optimal operation of a total H.323 solution.

System/network administrators have complete flexibility in defining zones. The network planner can use different criteria to architect an H.323 network to meet specific enterprise needs. For example, zones can be defined according to geographic locations (such as different branch locations) or in accordance with overlap of a physical network connection (such as a subnet on the floor of a building or a range of IP addresses), or by a functional (organizational) paradigm.

Identifying endpoints to a zone is done using IP addresses, alias names or phone numbers. A network planner can configure a gatekeeper to allow a specific set of endpoints into the zone and provide users with unique privileges. As a matter of fact, a gatekeeper can offer endpoints in a zone a variety of optional services including the following:

Use a “routed” call signaling model to route call signaling and control channels to the appropriate entities in the network.

Implement logic for granting/denying terminals, gateways, and MCUs access to the associated network assets (bandwidth, gateways, MCU, directory services, etc.). This is accomplished by monitoring all concurrent calls in a zone and enforcing network management policies for any new calls (sessions) a user may initiate.

The gatekeeper is the focal point for insertion of logic into the H.323 network. It can be configured and controlled remotely by third-party applications using http or SNMP protocols. A specific example is a call center ACD (automatic call distribution) application imposing call routing logic onto H.323 traffic via the gatekeeper.

3. Selecting the Right Video-Enabling Components

When the network upgrades described above are underway, it is time to select the video-specific elements of a successful network. In some cases the video endpoint and server components will be selected in advance by a service provider and offered in the form of a solution “bundle” for a monthly rate. In other situations it will be appropriate for a business to own and deploy its own hardware and software. In either case, a network engineer should evaluate and test interoperability between several types of video-enabling products:

- Networking elements for zone management (i.e., H.323 gatekeepers)
- Video-ready endpoint devices
- Multipoint control units
- Networking solutions for business-to-business connectivity

Selecting H.323 Gatekeepers

In many cases, a vendor has developed network solutions for use with a specific gatekeeper. The enterprise manager will need to evaluate the benefits of the hardware and software combination. In general, the rule of thumb is to request that the manufacturer of any equipment provide a list of the gatekeepers with which the product has been tested.

Regarding the performance of gatekeepers, network designers should be aware that if the gatekeeper resides in a multi- or general-purpose client or server endpoint, gatekeeper performance can become processor-limited if the terminal must execute other resource intensive tasks.

In addition, if the video network applications will be integrated with other business applications, an application programming interface will be important. For example, if a company wants to integrate video services into a enterprise portal, the HTML and XML tags to the gatekeeper must be supported by the gatekeeper.

If a distributed network architecture is envisioned or planned, the gatekeepers in unique zones must be easily managed as a group. Remote management and monitoring of gatekeepers is also important for scaling and supporting a growing video application service. Network managers should also be able to export data from any gatekeeper to produce reports or usage logs and for other business management functions.

In general, the selection of a gatekeeper is key to a successful deployment because it is central to the implementation of numerous services for the user and network manager.

Gatekeepers

Gatekeepers are often integrated with other voice and video network components but can also be purchased separately.

Even when integrated with a vendor’s hardware platform, a standards-compliant gatekeeper will register, manage and control H.323 entities from third party providers. This list of gatekeeper suppliers is limited to companies with expertise in video over IP.

Vendors

Cisco Systems
elemedia
ezenia!
FVC.COM
PictureTel
RADVision
Sorenson Vision
VCON

Selecting the Video Endpoints

In applications requiring low levels of interactivity, the user is often seated in an office where a multimedia-equipped personal computer and network connection are available. Today's desktop computers only need software "clients" to display the video and audio, and users can interact with the file on the server (or the speaker/presenter at their leisure) with a keyboard and mouse. The clients may be standalone applications or browser plug-ins.

In contrast, video endpoints for conferencing offer the user potentially more interactivity, and the settings in which conferences take place can vary widely. The video and audio capture and compression quality will have an impact on all the participants of a conference. A video session may involve a conference room, an executive office and a cubicle, for example. The video endpoint output may go to a computer monitor, VGA display or a television monitor (an NTSC display). There are multiple user interfaces to choose from: a PC–based application, a Web-based portal (with run-time extensions) or a handheld remote control for call initiation and other aspects of session management.

The ultimate choice of video endpoint design should be driven by the application requirements, usability, cost, interoperability needs and other business issues, such as the supplier's ability to furnish peripherals (e.g., for data conferencing), master agreements with systems integrators that carry a limited line of products, etc.

Once having weighed the unique requirements identified in the needs analysis and the options outline above, each location will have at least the following:

- a video input/capture device (camera) and video display
- compression and decompression device
- an audio capture and sound output device
- a network interface
- a user interface integrated with network protocol stack the permits the endpoint to place and respond to network requests

One of the tricky issues that designers should be aware of is that voice and videoconferencing endpoints will adjust to fluctuating network conditions differently. Some units adjust the data volume very aggressively in both directions while others adjust down rapidly but upward slowly. If the network does not present to the endpoint a somewhat stable amount of usable bandwidth, then those units that adjust aggressively upward present a poorer picture over the course of the conference. If there are just a few minor bandwidth problems during the conference, then a unit that generates as much video data as possible within the policy constraints will present a better picture. So the choice of endpoint depends on what type of network is available.

When selecting a voice or video-enabled endpoint for use in an IP environment, evaluators must ask about each manufacturer's traffic-shaping algorithms. Are the endpoints capable of monitoring network conditions and making adjustments? For example, do the encoders reduce the rate at which they introduce data on the network when they sense increasing network congestion? How does the application control the rate at which the product puts video onto the network? How does a receiving endpoint process out-of-order or duplicate packets it receives before decoding them?

Packet scheduling is a sender feature that buffers audio/video packets once they are compressed and then injects them into the network at a steady pace. This is effective in reducing congestion, but the downside to packet scheduling is that it can be computationally demanding. A video-encoding endpoint must have at least a 400-MHz processor in the CPU or use digital signal processors to manage the scheduling.

During a videoconference, packets arrive at an endpoint's network interface out-of-order due to the connectionless nature of IP networks. Without packet-ordering software to sort through and drop out-of-order packets, the end user will detect visual artifacts such as blocks of misplaced colors in the video or pops in the audio. Endpoints manage packet-arrival irregularities differently. Some applications are designed to freeze the video completely until the next full frame can be created free of artifacts. Others display frames with "ghosts" (like a trail produced by a rapid movement in a strobe light) of images from previous frames. Out-of-order packets also contribute to a lack of lip synchronization on the screen with audio.

End points will also differ in their approach to firewalls (see “security” below). They also have different degrees of compliance with the H.323 standard and could encounter interoperability problems in a mixed vendor network. The bottom line is that those evaluating video network systems will need to request and obtain a list of endpoints with which the network components have been certified or tested.

Selecting Multipoint Control Units

Videoconferences involving more than two endpoints must involve additional software or hardware resources. A multipoint control unit provides such functionality. The functionality relies on both multipoint processing and multipoint control software. Multipoint processors perform functions such as audio mixing and sense and process voice for source switching and transcoding (changing the data rate or format, or both, of video and audio to meet the requirements of individual endpoints in a call). The multipoint control software manages the distribution of these functions across one or multiple processors.

The selection of an MCU should consider if the application will benefit from continuous presence, the view of multiple participants on one display. For continuous presence, hardware transcoding drastically reduces bandwidth needs by composing a new video frame using the video data streams from multiple sources.

If the deployment is going to enable remote offices with their own MCU resources, the criteria for selection will prioritize lower-cost, low–port density devices with the ability to be controlled from a centralized operations center.

Vendors

H.323 MCUs can be software “only” (the supplier ships software for an off-the-shelf server operating system) or hardware-accelerated (in which case DSPs are optimized to accelerate MCU–specific processes such as transcoding and video mixing).

SW-only MCU providers:

CUseeMe Networks
ezenia!
PictureTel

HW-accelerated MCU providers:

Accord Networks
ezenia!
Lucent
RADVision

4. Business-to-Business Connectivity

The farther apart video participants are from one another, the greater the potential time savings and productivity enhancements associated with using video as an alternative to face-to-face communications. Today, a video-enabled conference room or desktop endpoint uses H.320 signaling and ISDN, end-to-end, for its conferencing needs.

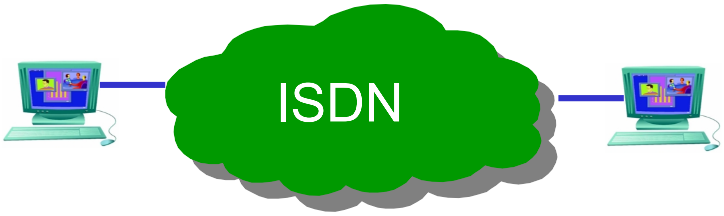


Figure 5. H.320-to-H.320 over ISDN

This is the most popular solution today, but cost and reach of ISDN prohibit ubiquitous deployment. If, for example, an engineer on the Ford Motor company network wants to communicate, collaborate and conference visually with a station in an independent manufacturing facility, IP would most likely be involved because the remote endpoints are connected to IP networks, not ISDN. While using an IP network makes endpoints and network equipment easier to manage in the enterprise, the options for business-to-business connectivity are far less clear.

One hybrid IP/ISDN–based solution is to have IP endpoints sharing access to ports on one or more H.323/H.320 gateways. This is in effect a pooling of centrally-managed hardware resources, most likely deployed in a telephony network closet. H.323 would be the protocol used for the IP leg to the ISDN interface (gateway). A gateway at the LAN/WAN interface would convert signaling from H.323 to H.320.

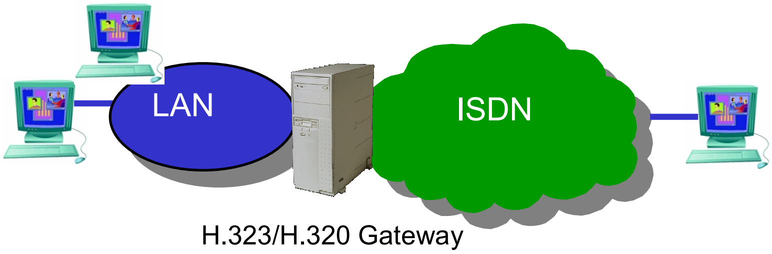


Figure 6. from H.323 (IP) to H.320 over ISDN

Video Network Gateways

Video network gateways convert video caller signaling and media from one network or protocol to another network environment on behalf of callers. Most solutions include *at least* two network interfaces and will support multiple simultaneous calls. An ISDN network to Ethernet network gateway is the most common configuration. Other configuration options include ATM–to–IP and IP–to–IP gateways.

Vendors

Accord Network
ezenia!
FVC.COM
RADVision

To have H.320/H.323 gateways support IP–to–IP video communications, the caller would require another gateway to convert the video back from H.320 format (ISDN) to

H.323 for transmission over IP to the remote user. Unfortunately, the delay introduced by having two video network gateways in each stream would dramatically degrade the quality of the video. H.321 [ISDN over ATM] is an attractive alternative because there is no need for translation of signaling or audio and video transcoding. Direct inward dialing (DID), a telephony feature, supports calls coming in from ISDN to IP/ATM through a gateway.

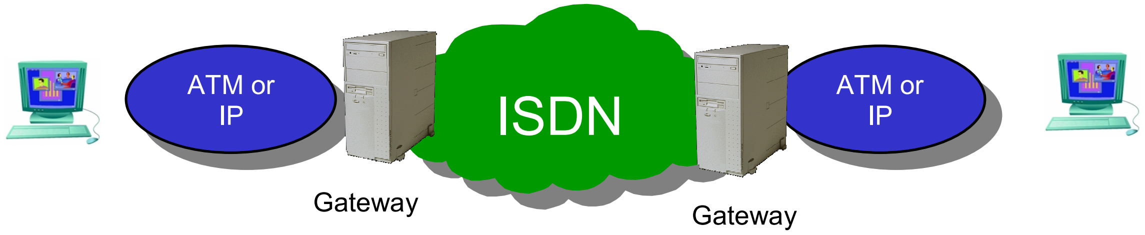


Figure 7. From H.323 (IP) to H.323 (IP) over ATM or ISDN

Another possible scenario is to have a separate physical network with a broadband connection, such as T1, DSL or a cable modem, and a non-secure connection to the public Internet. Provided the parties wanted to meet at a time of day when the backbones of the public Internet are not congested, this topology would probably work. However, it is not commercially attractive for two reasons: the lack of a standardized global addressing scheme and the fact that the endpoint could not be connect to the corporate network resources.

Addressing

For people to place telephone calls to others, they need to have a telephone number or a directory of names from which to obtain a corresponding telephone number. This is a standardized global addressing scheme. In data networking, the primary relationship is between the client and the server, as opposed to peer-to-peer relationships used in telephony. The client always “calls” the server for information or to change data on the server. These servers were primarily for internal use. Web servers, the exception to the rule of internally hosted data, are often hosted by an ISP or data center with special Web access/hosting services.

When the population of client computers drastically exceeded the Internet’s capacity for unique public IP addresses, the Internet’s architects hoped IP version 6 (IPv6) would overcome the problem. Unfortunately, the conversion to IPv6 failed to meet the urgent needs of the Internet. Network address translation (NAT) was widely implemented in enterprise networks to reduce the need for unique public IP addresses and to solve client-to-server addressing problems.

NAT allows networks to use private IP addresses that will be translated to public IP addresses when accessing the Internet. This is also an excellent way to secure a network from intrusion because a computer outside the NAT device can’t communicate directly with devices inside because the private addresses are not routable over the Internet.

However, because the private addresses are not routable outside the private network, NAT is not compatible with present H.323 dialing schemes. An H.323 call must be established with an endpoint or entity with a unique IP address, and that implies and includes a complete set of associated ports in specific address ranges, with which various TCP/IP and UDP sessions are established. With a 323-aware NAT present, a single H.323 endpoint can place calls out. However, it can never receive inbound calls.

A specially configured proxy/gatekeeper can work alongside a NAT server and take advantage of port-forwarding capabilities to make sure that the conferencing data reaches the intended recipient inside the network (supporting inbound calls).

The above solutions address situations in which a private IP address is involved. However, a global addressing plan, such as that in use for telephony, has yet to be standardized. Currently, when a terminal wants to place a call and it does not know the address of the party, it sends a query to the gatekeeper. If the said gatekeeper does not have the address for the called party, it sends out the same query to neighbor gatekeepers. One way to query multiple gatekeepers is to use multicast. This will only work in environments where multicast is deployed and is not considered secure because anyone can receive and respond to the query. The other way to have gatekeepers communicate is for each gatekeeper to maintain a database of gatekeepers known to it. This solution is not scalable.

The lack of a coordinated global addressing scheme means that two or more entities could adopt the same video number.

Security

Virtual private network is one of the best ways to address security while allowing video calls to be placed inside an enterprise with multiple sites, but is not appropriate for business-to-business communications.

When an H.323 call is set up, as is routine in telephony today, the call recipient's application or server must accept an *incoming TCP session on a random port from an unknown IP address*. Allowing an unknown IP address access to an enterprise network is a high-risk proposition, one that firewalls are in place to prevent.

Vendors

Proprietary solutions are available to solve the addressing and security issues in 2-way video over IP. Cisco's PIX, combined with NAT and the proxy gatekeeper, offers an integrated solution for networks using Cisco routers. Alternatives, such as Sorenson Vision's Glasses proxy gatekeeper, offer a solution for non-Cisco networks. Both these systems typically rely on communicating with the NAT or proxy server from a directory that specifies the NAT's IP address *and* the receiver's alias (usually an e-mail address).

Accord Networks' VGC-20 is an integrated hardware/software solution to address both the addressing and firewall problems with an H.323-to-H.323 gateway that initiates a session with endpoints both on and off the secure network at the same time.

Though it does not address common security issues, Polycom's Global Management System offers a centrally managed Web-based directory service solution. While greater diagnostics and reporting are available for Polycom endpoints, an IP address or ISDN number can be used when populating the directory with third-party endpoints.

An H.323 video call involves multiple communication sessions. There are separate video and audio UDP streams and control signaling that transmit both TCP and UDP packets. While UDP is a connectionless communication, not generally regarded as a security risk, TCP sessions are connection-oriented and represent a potential security risk. There are multiple ways to maintain security while also permitting parties on different secure networks to call one another:

- IP filters with gatekeeper-directed calling
- H.323 proxy gatekeeper

Gatekeeper-directed calling is basically gatekeeper-directed routing that is always initiated by the gatekeeper. In this situation, the call gets initiated by a "known and trusted" address using all standards-compliant signaling.

A known and trusted MCU can also be used to initiate separate video calls between different customers. This also provides a secure mechanism for exchanging video between customer sites without using direct IP traffic between the two caller networks. The benefit of using an MCU to bridge the call is that the calling endpoint can initiate the call to the MCU without security risk. The down side is that every endpoint must "take an action" to complete a call, a situation that is not uncommon in video networking today.

Setting up calls using H.323 proxy gatekeepers, as discussed above under addressing, is also secure because the receiver's IP address is unknown to the caller. Therefore, the security is never breached during the session.

Outsourcing Business-to-Business Connectivity and Applications

As discussed above, video, as a component of a complete communications solution, has suffered from a number of inhibitors. Recently increased access to greater bandwidth networks, new endpoint and network technologies and the Web are all contributing to decrease the barriers to adoption of video.

New packaging of hardware, software and services in value bundles will accelerate adoption by offering higher ease of use, greater reliability and lower costs structures. To build and offer successful managed video services, companies of different types will need to have partnerships. The partnerships will impact the three levels of a new network and application model: the customer layer, the network infrastructure layer and the video services layer. Using this model, enterprise customers will outsource the management of their video services. One of the numerous benefits of outsourcing business to business and intra business video networks is that the customer's sessions will be able to rely on professionally managed resources. These networks and applications will have a consistent QoS as designated in the user's profile, regardless of the caller's destination. Furthermore, the enterprise customer will experience cost savings from using the service provider's infrastructure.

The long-term solution to the question of inter-domain or inter-business videoconferencing will be resolved through new devices in the network service providers' infrastructure that will leverage Signaling System 7 (SS7) with relational databases and supplementary network services commonly found in the public telephone network today, but otherwise absent in the IP world. All these services will emerge transparently to the user and will be offered as an option to the enterprise network manager.

5. Refinements and Getting to Full-Scale Deployment

Once the pilot phase of a video network deployment has proven its worth, the network users should plan for a refinement of services over time. Some new services may exceed an IT group’s capabilities, in which case the company will turn to a service provider or systems integrator with experience in deploying video on IP networks for help.

As the ability to use networked multimedia reaches more companies, the entire community will grow and, like a snowball, quickly reach critical mass. Critical mass in business will also drive and be driven by the ever-escalating size of the consumer, networked, multimedia-ready audience.